As per CBCS syllabus of UGC for Indian Universities

GROUP THEORY & LINEAR ALGEBRA II

Theory, Problems & Solutions

RANEN BHATTACHARYYA
Associate Professor, Department of mathematics
Bijoy Krishna Girls' College, Howrah

ACADEMIC PUBLISHERS
5A, Bhawani Dutta Lane, KOLKATA-700 073
E-mail : contact@academicpublishers.in
Website : www.academicpublishers.in

© Reserved by the author

First edition 2021
Reprint 2022

ISBN : 978-93-87162-70-9

Price : Rupees one hundred and seventy five only.

Dedicated to
my beloved students

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise, without the prior permission of the copyright holder.

Published by B. K. Dhur of Academic Publishers, 5A, Bhawani Dutta Lane, Kolkata-700 073, laser type setting by Studio Michaelangelo, 5A, Bhawani Dutta Lane, Kolkata-700 073 and printed at Rajendra Offset, 11 Panchanan Ghosh Lane, Kolkata-700 009.

Preface to the first edition

This book is a continuation of my previous books Group Theory 1 and Ring Theory and Linear Algebra. Students are probably about to begin their second exposure to Group Theory and Linear Algebra. Unlike their first brush with the subject, which probably emphasized Groups, Subgroups, Cyclic Groups, Normal subgroups, Isomorphisms, Vector spaces, Basis, Matrices etc. we will focus on Automorphisms, Product of Groups, Inner Product Spaces, Linear Maps, Operators, Canonical forms etc. These terms will be defined later, so don't worry if you don't know what they mean. The main goal of this book is to make those concepts palatable. The key point is that students are about to immerse themselves in serious mathematics, with an emphasis on their attaining a deep understanding of the definitions, theorems and proofs.

I wish to express my gratitude to my teachers who have inspired me. I am also thankful to my students also because I have learnt a lot from them.

Many thanks are due also to Mr Bimal kumar Dhur, Prof. Subhankar Dhur and Mr Dipankar Dhur of Academic Publishers for their continuous help and cooperation in this endeavour.

Lastly, I would like to thank Mrs Susmita (Sumon) Bhattacharya, my wife, without whose constant support this work would never been possible.

I would greatly appreciate hearing about any errors in this book, even minor ones. I welcome your suggestions for improvements, even tiny ones. Please feel free to contact me by email at **ranenpersonal@gmail.com**

Have fun !

Kolkata
December, 2020

Ranen Bhattacharyya

Contents

Unit-1

Group Theory

1.1 AUTOMORPHISM

It is presumed that students are already aware of groups and group isomorphism. But for recapitulation we wish to offer following definitions.

Definition: Let $(G, o), (G', *)$ be two groups. A mapping $f: G \to G'$ is called a **homomorphism** if $f(aob) = f(a) * f(b)\ \forall a, b \in G$.

Definition: Let $(G, o), (G', *)$ be two groups. A homomorphism $f: G \to G'$ is

(a) **Epimorphism** if f is onto i.e. $f(G) = G'$

(b) **Monomorphism** if f is injective.

(c) **Isomorphism** if f is bijective.

Example: $(\mathbb{Z}, +)$ is isomorphic to $(2\mathbb{Z}, +)$.

Define $f: \mathbb{Z} \to 2\mathbb{Z}$ by $f(n) = 2n$ for $n \in \mathbb{Z}$.

Let $f(n) = f(m)$. Then $2n = 2m$ i.e. $n = m$. So f is injective.

If $y \in 2\mathbb{Z}$ then $y = 2k, k \in \mathbb{Z}$. Hence $f(k) = 2k = y$. So f is onto.

For $m, n \in \mathbb{Z}$, $f(m+n) = 2(m+n) = 2m + 2n = f(m) + f(n)$. So f is a homomorphism. Hence f is an isomorphism.

It is clear from above that a mapping $f : G \to G'$, where G and G' are groups, is an isomorphism if

$(i)\ f(ab) = f(a)f(b)$

$(ii)\ f\ is\ one - one$

$(iii)\ f(G) = G'$

Here operations are taken as multiplication. But what happens if we take $G' = G$, that is, what happens if we consider an isomorphism from a group G onto itself ?

Definition. An isomorphism from a group G onto itself is called an **automorphism of G.**

Examples :

(i) For any group G, the identity mapping $i : G \to G$ by $i(x) = x$ is an automorphism of G.

(ii) Let $\mathbb{R}^+$ be the set of all positive reals. Then we know that $\mathbb{R}^+$ forms a group with respect to multiplication. Define $f : \mathbb{R}^+ \to \mathbb{R}^+$ by $f(x) = x^2$.

For $x, y \in \mathbb{R}^+$, we have, $f(xy) = (xy)^2 = x^2y^2 = f(x)f(y)$. Therefore, f is a homomorphism.

Again, $f(x) = f(y) => x^2 = y^2 => x = y$ as $x > 0, y > 0$. So, f is injective.

For $y \in \mathbb{R}^+$, we have, $\sqrt{y} \in \mathbb{R}^+$ and $f(\sqrt{y}) = (\sqrt{y})^2 = y$. Therefore, f is onto.

Hence, f is an automorphism.

(iii) We know that $(\mathbb{C}, +)$, $\mathbb{C}$ being the set of all complex numbers, is a group. Define $f : \mathbb{C} \to \mathbb{C}$ by $f(z) = \bar{z}$, that is, $f(a+ib) = a - ib$ where a, b are reals. Then for any $z_1, z_2 \in \mathbb{C}$,

$$f(z_1 + z_2) = \overline{z_1 + z_2} = \bar{z_1} + \bar{z_2} = f(z_1) + f(z_2)$$

Let $z, w \in \mathbb{C}$ where $z = a + ib, w = c + id,\ a, b, c, d \in \mathbb{R}$. Now,

$$f(z) = f(w) => \bar{z} = \bar{w} => a - ib = c - id => a = c, b = d$$

Thus, $f(z) = f(w)$ implies $z = w$, that is, f is injective.

Let $z \in \mathbb{C}$. Then $\bar{z} \in \mathbb{C}$ and $f(\bar{z}) = \bar{\bar{z}} = z$. Thus f is onto.

Hence, f is an automorphism of $\mathbb{C}$.

If $\mathbb{C}^*$ denotes the set of non-zero complex numbers then we know that $\mathbb{C}^*$ forms a group with respect to multiplication. In the same way, it is easy to show that $f : \mathbb{C}^* \to \mathbb{C}^*$ defined by $f(z) = \bar{z}$ is an automorphism of $\mathbb{C}^*$.

(iv) If we take $G = \mathbb{R}^2 = \{(a, b) : a, b \in \mathbb{R}\}$. Then $(G, +)$ is a group if '+' is defined as

$$(a, b) + (c, d) = (a + c, b + d)$$

Let us define $f : G \to G$ by $f(a, b) = (b, a)$. Then,

$$f((a, b) + (c, d)) = f((a + c, b + d)) = (b + d, a + c)$$
$$= (b, a) + (d, c) = f(a, b) + f(c, d)$$

So, f is a homomorphism. Now,

$$f(a, b) = f(c, d) => (b, a) = (d, c) => b = d, a = c => (a, b) = (c, d)$$

Therefore, f is injective.

For $(a, b) \in G$, there exists $(b, a) \in G$ such that $f(b, a) = (a, b)$. Thus f is onto.

Hence, f is an automorphism.

(v) Let G be a group and let $a \in G$.

Define a function $\phi_a : G \to G$ by $\phi_a(x) = axa^{-1}$.

Now for $x, y \in G$, we have,

$$\phi_a(xy) = axya^{-1} = axa^{-1}aya^{-1} = \phi_a(x)\phi_a(y)$$

So, ϕ_a is a homomorphism.

$\phi_a(x) = \phi_a(y) => axa^{-1} = aya^{-1} => x = y$ (by cancellation laws in G)

Therefore, ϕ_a is one-one.

For $y \in G$, we have, $a^{-1}ya \in G$ and $\phi_a(a^{-1}ya) = a(a^{-1}ya)a^{-1} = y$.

Thus, ϕ_a is onto.

Hence, ϕ_a is an automorphism. This ϕ_a is of special interest as clear from the following definition.

Definition.

Let G be a group and let $a \in G$. The function $\phi_a : G \to G$ defined by $\phi_a(x) = axa^{-1}$ for all $x \in G$ is called the **inner automorphism corresponding to a** or the **inner automorphism of G induced by a.**

Let G be a group. Let I be the identity mapping of G, that is, $I(x) = x, \forall x \in G$.

Then I is an automorphism of G. Let $\mathcal{A}(G)$ or $Aut(G)$ be the set of all automorphisms of G, that is,

$$Aut(G) = \mathcal{A}(G) = \{f : G \to G \mid f \text{ is an automorphism of } G\}$$

Clearly, $\mathcal{A}(G)$ is a subset of $S(G)$, the set of all permutations of G. Define product in $\mathcal{A}(G)$ as the compositions of mappings, that is, for $f, g \in \mathcal{A}(G)$, $(fg)(x) = f(g(x)) \forall x \in G$ which is defined in $S(G)$. We want to show that $\mathcal{A}(G)$ is a subgroup of $S(G)$.

Since, $f, g \in \mathcal{A}(G)$, we have, $f(xy) = f(x)f(y)$ and $g(xy) = g(x)g(y)$ for all $x, y \in G$ and f, g are bijective. Since, composition of two bijective mappings is bijective, we have, fg is bijective.

For $x, y \in G$, and for $f, g \in Aut(G)$, using homomorphism properties of f and g, we have,

$$fg(xy) = f(g(xy)) = f(g(x)g(y)) = f(g(x))f(g(y)) = fg(x)fg(y)$$

Thus, fg is a homomorphism and hence is an automorphism of G.

Therefore, $f, g \in \mathcal{A}(G) => fg \in \mathcal{A}(G)$.

Now, it is enough to show that $f \in \mathcal{A}(G) => f^{-1} \in \mathcal{A}(G)$. Since, f is bijective, f^{-1} exists and f^{-1} is also bijective.

Let $x, y \in G$. Then

$$f(f^{-1}(x)f^{-1}(y)) = f(f^{-1}(x))f(f^{-1}(y)) = (Ix)(Iy) = xy$$

If $(\phi(a))^m = e$ for some integer m with $0 < m < n$, then we have,

$$\phi(a^m) = e = \phi(e)$$

Since, ϕ is one-one, we have, $a^m = e$, which contradicts that $o(a) = n$ as $0 < m < n$.

Therefore, $(\phi(a))^m \neq e$ for any integer m with $0 < m < n$.

Hence, $o(\phi(a)) = n = o(a)$.

1.2 AUTOMORPHISM GROUPS OF FINITE AND INFINITE CYCLIC GROUPS

Let G be a finite cyclic group of order r, that is, let $G = < a >$ where $a^r = e$. Suppose T is an automorphism of G. Let $g \in G$. Then $g = a^k$ for some $k \in \mathbb{Z}$. Thus,

$$T(g) = T(a^k) = (T(a))^k [\,as\ T\ is\ automorphism]$$

Hence, $T(g)$ is completely determined for any $g \in G$, if $T(a)$ is known.

Now it is shown in theorem 1.1.2 that $o(T(a)) = o(a) = r$.

Since, $T(a) \in G$ and $G =< a >$, we have,

$T(a) = a^t$ for some t with $0 < t < r$.

So, $o(T(a)) = o(a) => o(a^t) = r$, which shows that $\gcd(t, r) = 1$.

Hence, for each automorphism T of G, we get an integer t which is less than r and prime to r. Thus, $Aut(G)$, the group automorphism of G, is in one-to-one correspondence with the group U_r of integers less than r and relatively prime to r under multiplication modulo r.

Let us rename the elements of $Aut(G)$ as T_i where $T_i(a) = a^i$ for $0 < i < r$ and $\gcd(i, r) = 1$.

Now, $T_iT_j(a) = T_i(a^j) = a^{ij} = T_{ij}(a)$. Therefore, $T_iT_j = T_{ij}$.

Define $\phi : U_r \to Aut(G)$ by $\phi(i) = T_i$. Then

$$\phi(ij) = T_{ij} = T_iT_j = \phi(i)\phi(j)$$

So, ϕ is a homorphism. It is shown that ϕ is bijective. Hence ϕ is an isomorphism.

Hence, $Aut(G) \cong U_r$.

What happens if G is an infinite cyclic group?

Let $G = \{a^k : k \in \mathbb{Z}\}$ be an infinite cyclic group generated by a. Here, $a^k = e$ if and only if $k = 0$.

Let T be an automorphism of G. Then $T(a) \in G =< a >$. So there exists $t \in \mathbb{Z}$ such that $T(a) = a^t$.

Since T is an automorphism of G, we have, $T(G) = G$. So, there exists $g \in G$ such that $T(g) = a$.

Again, $g \in G$ implies that there exists $i \in \mathbb{Z}$ such that $g = a^i$

Thus, $a = T(g) = T(a^i) = (T(a))^i = (a^t)^i = a^{ti}$

which shows that $a^{ti-1} = e$, that is, $ti - 1 = 0$, that is. $ti = 1$.

Since, t and i both are integers, we have, two possibilities : either $t = 1, i = 1$ or $t = -1, i = -1$.

If $t = 1$, we have, $T(a) = a$. If $x \in G$ then $x = a^k$ for some $k \in \mathbb{Z}$. Thus, we have,

$$T(x) = T(a^k) = (T(a))^k = a^k = x$$

That is, $T(x) = x$ for all $x \in G$ which shows that T is identity automorphism.

If $t = -1$, we have, $T(a) = a^{-1}$. Thus, for $x \in G$, we have,

$$T(x) = T(a^k) = (T(a))^k = (a^{-1})^k = (a^k)^{-1} = x^{-1}$$

That is, $T(x) = x^{-1}$ for all $x \in G$.

Hence, if G is an infinite cyclic group, then there are only two automorphisms of G, one is identity automorphism and other takes $g \to g^{-1}$ for all $g \in G$, in other words, $Aut(G)$ is isomorphic to a cyclic group of order 2.

Solved Problems :

1. **Let G be a group, H a subgroup of G, T an automorphism of G. Let $T(H) = \{T(h) : h \in H\}$. Prove that $T(H)$ is a subgroup of G.**

Solution. By the problem, $T : G \to G$ is a homomorphism and bijective.

Now, $e \in H => T(e) \in T(H) => T(H) \neq \emptyset$.

Let $k_1, k_2 \in T(H)$. Then there exist $h_1, h_2 \in H$ such that $T(h_1) = k_1$, $T(h_2) = k_2$. Now,

$$k_1k_2^{-1} = T(h_1)T(h_2)^{-1} = T(h_1)T(h_2^{-1}) = T(h_1h_2^{-1})[\,as\ T\ is\ homomorphism]$$

Since, H is a subgroup of G, we have, $h_1, h_2 \in H => h_1h_2^{-1} \in H$. Thus, $T(h_1h_2^{-1}) \in T(H)$.

Therefore, $k_1k_2^{-1} = T(h_1h_2^{-1}) \in T(H)$.

Hence, $T(H)$ is a subgroup of G.

2. **Let G be a group. Prove that the mapping $f(g) = g^{-1}$ for all $g \in G$ is an automorphism if and only if G is abelian.**

Solution. We first suppose that $f : G \to G$ given by $f(g) = g^{-1}$ is an automorphism.

Let $a, b \in G$. Then $f(a) = a^{-1}, f(b) = b^{-1}$.

Again, $f(ab) = f(a)f(b)$

$$=> (ab)^{-1} = a^{-1}b^{-1} = (ba)^{-1} => ab = ba$$

Therefore, G is abelian.

Conversely, let G be abelian. Let $a, b \in G$. Then $ab = ba$. Now,

$$f(ab) = (ab)^{-1} = (ba)^{-1} = a^{-1}b^{-1} = f(a)f(b)$$

So, f is a homomorphism.

Let $g, h \in G$ such that $f(g) = f(h)$, i.e., $g^{-1} = h^{-1}$, i.e., $g = h$.

So, f is injective.

For any $g \in G$, there exists $g^{-1} \in G$ such that $f(g^{-1}) = (g^{-1})^{-1} = g$. So, f is onto.

Hence, f is an automorphism of G.

3. Let G be a group, T an automorphism of G, N a normal subgroup of G. Prove that $T(N)$ is a normal subgroup of G.

Solution. . By the problem, $T : G \to G$ is a homomorphism and bijective.

Now, $e \in N => T(e) \in T(N) => T(N) \neq \emptyset$.

Let $k_1, k_2 \in T(N)$. Then there exist $h_1, h_2 \in N$ such that $T(h_1) = k_1$, $T(h_2) = k_2$. Now,

$$k_1k_2^{-1} = T(h_1)T(h_2)^{-1} = T(h_1)T(h_2^{-1}) = T(h_1h_2^{-1})[\,as\ T\ is\ homomorphism]$$

Since, N is a subgroup of G, we have, $h_1, h_2 \in N => h_1h_2^{-1} \in N$. Thus, $T(h_1h_2^{-1}) \in T(N)$.

Therefore, $k_1k_2^{-1} = T(h_1h_2^{-1}) \in T(N)$.

Hence, $T(N)$ is a subgroup of G.

Let $g' \in G$ and $k \in T(N)$. As $T : G \to G$ is onto, there exists $g \in G$ such that $T(g) = g'$.

Again, $k \in T(N)$ implies that there exists $h \in N$ such that $T(h) = k$. Thus,

$$g'kg'^{-1} = T(g)T(h)\big(T(g)\big)^{-1} = T(g)T(h)T(g^{-1}) = T(ghg^{-1})$$

Since N is a normal subgroup of G, we have, $g \in G, h \in N => ghg^{-1} \in N$. Hence,

$$g'kg'^{-1} = T(ghg^{-1}) \in T(N)$$

Therefore, $T(N)$ is a normal subgroup of G.

4. Let G be a group of order 4, $G = \{e, a, b, ab\}$, $a^2 = b^2 = e$, $ab = ba$. Determine $Aut(G)$.

Solution. By the problem, we see that order of a and b are 2. Now,

$$(ab)^2 = abab = a^2b^2(\,as\ ab = ba) = e$$

and $ab \neq e$ otherwise, $a = b^{-1} = b$ which is not the case. Thus, order of ab is 2.

It is clear that proper subgroups of G are given by $\{e, a\}, \{e, b\}, \{e, ab\}$.

If T is an automorphism of G, then $T(e) = e$. Again, order of $T(a)$ is 2 as order of a is 2. So, there are three possibilities of $T(a)$, viz. a, b, ab.

Since, T is a homomorphism, we have $T(a)T(b) = T(ab)$.

If $b \neq e$, we have, the order of $T(b)$ is 2 as the order of b is 2. So, again there are three possibilities of $T(b)$ viz. a, b, ab. But one of these three members is already associated with $T(a)$. So, two possibilities remain only.

Hence, there are only $3 \times 2 = 6$ automorphisms of G. Thus $Aut(G)$ is given by

$$Aut(G) = \left\{\begin{matrix} \begin{pmatrix} e & a & b & ab \\ e & a & b & ab \end{pmatrix}, \begin{pmatrix} e & a & b & ab \\ e & b & a & ab \end{pmatrix}, \begin{pmatrix} e & a & b & ab \\ e & a & ab & b \end{pmatrix}, \\ \begin{pmatrix} e & a & b & ab \\ e & ab & b & a \end{pmatrix}, \begin{pmatrix} e & a & b & ab \\ e & b & ab & b \end{pmatrix}, \begin{pmatrix} e & a & b & ab \\ e & a & b & ab \end{pmatrix} \end{matrix}\right\}$$

5. Let G be a finite group, T an automorphism of G with the property that $T(x) = x$ for $x \in G$ if and only if $x = e$. Prove that every $g \in G$ can be represented as $g = x^{-1}T(x)$ for some $x \in G$.

Solution. Let us define a map $f : G \to G$ by $f(x) = x^{-1}T(x)$. Now,

$$f(a) = f(b) => a^{-1}T(a) = b^{-1}T(b)$$

$$=> T(a)\big(T(b)\big)^{-1} = ab^{-1}$$

$$=> T(ab^{-1}) = ab^{-1}$$

It is given that, $T(x) = x$ iff $x = e$ for all $x \in G$. Hence, $T(ab^{-1}) = ab^{-1} =>$ $ab^{-1} = e$ i.e. $a = b$.

Thus, $f(a) = f(b) => a = b$ which proves that f is injective. Since, G is finite, f is onto.

Therefore, for $g \in G$, there exists $x \in G$ such that $f(x) = g$ i.e. $x^{-1}T(x) = g$.

6. Let G be a finite group and T an automorphism of G with the property that $T(x) = x$ if and only if $x = e$. Suppose further that $T^2 = I$. Show that, G is abelian.

Solution. Let us define a map $f : G \to G$ by $f(x) = x^{-1}T(x)$.

Now, $f(a) = f(b) \Rightarrow a^{-1}T(a) = b^{-1}T(b)$

$$\Rightarrow T(a)(T(b))^{-1} = ab^{-1} \Rightarrow T(ab^{-1}) = ab^{-1}$$

It is given that, $T(x) = x$ iff $x = e$ for all $x \in G$. Hence, $T(ab^{-1}) = ab^{-1} \Rightarrow ab^{-1} = e$ i.e. $a = b$.

Thus, $f(a) = f(b) \Rightarrow a = b$ which proves that f is injective. Since, G is finite, f is onto.

Therefore, for $g \in G$, there exists $x \in G$ such that $f(x) = g$ i.e. $x^{-1}T(x) = g$. Now,

$$T(g) = T(x^{-1}T(x)) = T(x^{-1})TT(x) = (T(x))^{-1}x \text{ (as } T^2 = I)$$

$$= (x^{-1}T(x))^{-1} = g^{-1}$$

Thus, $T(g) = g^{-1}\ \forall g \in G$. So, for $a, b \in G$,

$$T(ab) = (ab)^{-1}$$

Again, $T(ab) = T(a)T(b) = a^{-1}b^{-1} = (ba)^{-1}$

Hence, $(ab)^{-1} = (ba)^{-1}$, i.e. $ab = ba$.

Therefore, G is abelian.

7. Show that $Aut\ (Z_n) \simeq U_n$.

Solution. Since Z_n is a cyclic group of order n, the result follows from article 1.2.

8. Find two groups G and H such that G and H are not isomorphic but $Aut(G) \approx Aut(H)$.

Solution. Let $G = (\mathbb{Z}_2, +), H = (\{0\}, +)$. Clearly, G and H are not isomorphic.

There is only one automorphism of G given by $f(1) = 1$.

Thus, $Aut(G) \approx Aut(H)$.

9. If a group G is isomorphic to H, prove that $Aut(G)$ is isomorphic to $Aut(H)$.

Solution. Let G and H be two groups and $\phi : G \to H$ be an isomorphism.

We define, $\beta : Aut(G) \to Aut(H)$ by $\beta(f) = \phi f \phi^{-1}$.

We first show that, $\beta(f + g) = \beta(f) + \beta(g)$.

Let $b \in H$. Then $\phi f \phi^{-1}(b) = \phi f(a)$ [where $\phi(a) = b, a \in G$]. Similarly, $\phi g \phi^{-1} = \phi g(a)$

$$\phi(f+g)\phi^{-1}(b) = \phi(f+g)(a) = \phi(f(a) + g(a)) = \phi f(a) + \phi g(a)$$
$$= \phi f \phi^{-1}(b) + \phi g \phi^{-1}(b)$$

Hence, $\phi(f+g)\phi^{-1} = \phi f \phi^{-1} + \phi g \phi^{-1}$,

i.e. $\beta(f + g) = \beta(f) + \beta(g)$

So, β is a homomorphism.

Let $f, g \in Aut(G)$ such that $\beta(f) = \beta(g)$, that is $\phi f \phi^{-1} = \phi g \phi^{-1}$

Let $x \in G$. Then $\phi(x) = y \in H$ and $\phi^{-1}(y) = x$ as ϕ is an isomorphism.

Now, $\phi f \phi^{-1}(y) = \phi g \phi^{-1}(y) \Rightarrow \phi f(x) = \phi g(x)$

Therefore, $f(x) = g(x), \forall x \in G$, as ϕ is one-one. Thus, $f = g$.

Hence, β is injective.

Let $h \in Aut(H)$. Then $\phi^{-1}h\phi \in Aut(G)$. Now,

$$\beta(\phi^{-1}h\phi) = \phi(\phi^{-1}h\phi)\phi^{-1} = h$$

Therefore, β is onto.

Hence, β is an isomorphism, in other words, $Aut(G) \approx Aut(H)$.

10. Let G be a group and Z the center of G. If T is any automorphism of G, prove that $T(Z) \subset Z$.

Solution. By the problem, $Z = \{x \in G : xg = gx, \forall g \in G\}$

Let $z' \in T(Z)$. Then there exists $z \in Z$ such that $T(z) = z'$.

Therefore, $zg = gz$ for all $g \in G$.

So, $T(zg) = T(gz)$, i.e., $T(z)T(g) = T(g)T(z)$, i.e., $z'T(g) = T(g)z'$ for all $g \in G$.

Since, $T : G \to G$ is onto, every element g' of G can be written as $T(g)$ for some $g \in G$. Thus,

$$z'g' = g'z', \quad \forall g' \in G$$

So, $z' \in Z$. Hence, $T(Z) \subset Z$.

11. Let G be a group and let $g \in G$. If $z \in Z(G)$, $Z(G)$ being the centre of G, show that the inner automorphism induced by g is the same as the inner automorphism induced by zg (that is, that the mappings ϕ_g and ϕ_{zg} are equal).

Solution. Let $x \in G$. Then $\phi_g(x) = gxg^{-1}$. Now,

$$\phi_{zg}(x) = zgx(zg)^{-1} = zgxg^{-1}z^{-1}$$
$$= (gxg^{-1})zz^{-1} [\text{as } z \in Z(G) \text{ and } gxg^{-1} \in G]$$

i.e. $\phi_{zg}(x) = gxg^{-1} = \phi_g(x),\ \forall x \in G$ and $dom\ \phi_g = dom\ \phi_{zg} = G$.

Hence, $\phi_g = \phi_{zg}$ ■

12. If g and h are elements from a group, prove that $\phi_g\phi_h = \phi_{gh}$.

Solution. Let G be a group and for $a \in G$, let ϕ_a be the inner automorphism induced by a. Then for $g, h \in G$, we have, $\phi_g(x) = gxg^{-1}$, $\phi_h(x) = hxh^{-1}$ and $\phi_{gh}(x) = ghx(gh)^{-1}$ for all $x \in G$.

Let $x \in G$. Then

$$\phi_g\phi_h(x) = \phi_g(hxh^{-1}) = g(hxh^{-1})g^{-1} = (gh)x(gh)^{-1} = \phi_{gh}(x)$$

Hence, $\phi_g\phi_h = \phi_{gh}$

13. Let G be a group and T an automorphism of G. If, for $a \in G, N(a) = \{x \in G : xa = ax\}$, prove that $N(T(a)) = T(N(a))$.

Solution. Here $N(T(a)) = \{g' \in G : g'T(a) = T(a)g'\}$.

Since $T : G \to G$ is onto there exists $g \in G$ such that $T(g) = g'$. Thus,

$$N(T(a)) = \{T(g) : g \in G \text{ and } T(g)T(a) = T(a)T(g)\}$$
$$= \{T(g) : g \in G \text{ and } T(ga) = T(ag)\}$$
$$= \{T(g) : g \in G \text{ and } ga = ag\} \text{ (as } T \text{ is one} - \text{one)}$$
$$= \{T(g) : g \in G \text{ and } g \in N(a)\} = T(N(a))$$

Hence, $N(T(a)) = T(N(a))$.

Exercise

Let G be a group.

1. Show that all automorphisms of G form a group under function composition.
2. Let ϕ be an automorphism of a group G. Prove that $H = \{x \in G : \phi(x) = x\}$ is a subgroup of G.
3. If $\sigma \in Aut\ (G)$, and ϕ_g is a conjugation by g, then prove that $\sigma\phi_g\sigma^{-1} = \phi_{\sigma(g)}$.
4. If G is an abelian group, then prove that the map $f : G \to G$ by $f(g) = g^{-1}$ is an automorphism.
5. Show that there are only two automorphisms of the group Z_6.
6. Show that $o\left(Aut\ \mathbb{Z}_p\right) = p - 1$ where p is prime.
7. How many automorphisms has a cyclic group of order ? of order pq ? (p, q distinct primes).

8. Show that $Aut(\mathbb{Z}) \simeq \mathbb{Z}_2$.
9. Prove that $Inn(S_3) \simeq S_3 \simeq Aut\ (S_3)$.
10. If G be a cyclic group of order n and ϕ be the Euler ϕ-function. Prove that $o(AutG) = \phi(n)$.
11. Show that $Inn\ (G)$ is a normal subgroup of $Aut\ (G)$.
12. Exhibit an automorphism of $\mathbb{Z}_6$ that is not an inner automorphism.
13. Prove that an element g of a group G induces the inner automorphism identity if and only if it is in the centre.

1.3 EXTERNAL DIRECT PRODUCT

Suppose two groups are given. Can we form a larger group with the help of given groups? Let's try.

Let (G, o) and $(H, *)$ be two groups. We consider the product

$G \times H = \{(g, h) : g \in G, h \in H\}$.

Let us define an operation '.' on $G \times H$ by

$$(g_1, h_1).(g_2, h_2) = (g_1 o g_2,\ h_1 * h_2)$$

Clearly,'.'is closed as $g_1, g_2 \in G \Rightarrow g_1 o g_2 \in G$ and $h_1, h_2 \in H \Rightarrow h_1 * h_2 \in H$. Thus

$$(g_1, h_1).(g_2, h_2) \in G \times H \Rightarrow (g_1 o g_2, h_1 * h_2) \in G \times H$$

Let $(g_1, h_1), (g_2, h_2), (g_3, h_3) \in G \times H$. Then

$$[(g_1, h_1).(g_2, h_2)].(g_3, h_3) = (g_1 o g_2, h_1 * h_2).(g_3, h_3)$$
$$= (g_1 o g_2 o g_3, h_1 * h_2 * h_3)$$

and $(g_1, h_1).[(g_2, h_2).(g_3, h_3)] = (g_1, h_1).(g_2 o g_3, h_2 * h_3)$

$$= (g_1 o g_2 o g_3, h_1 * h_2 * h_3)$$

Hence,

$$[(g_1, h_1).(g_2, h_2)].(g_3, h_3) = (g_1, h_1).[(g_2, h_2).(g_3, h_3)]$$

which shows that '.' is associative.

Clearly, $(e_G, e_H) \in G \times H$ where e_G and e_H are the identities of G and H respectively. Then

$$(g, h).(e_G, e_H) = (g o e_G, h * e_H) = (g, h) = (e_G o g, e_H * h) = (e_G, e_H).(g, h)$$

So, (e_G, e_H) acts as an identity element in $G \times H$.

Let $(g, h) \in G \times H$.

Now, $g \in G \Rightarrow g^{-1} \in G$ and $h \in H \Rightarrow h^{-1} \in H$.

Therefore, $(g^{-1}, h^{-1}) \in G \times H$.

So, $(g^{-1}, h^{-1}) \in G \times H$.

Then $(g,h).(g^{-1},h^{-1}) = (gog^{-1}, h * h^{-1}) = (e_G, e_H)$
and $(g^{-1},h^{-1}).(g,h) = (g^{-1}og, h^{-1} * h) = (e_G, e_H)$

Thus inverse of (g,h) is (g^{-1},h^{-1}) and it belongs to $G \times H$.

Hence, $(G \times H, .)$ is a group, known as *external direct product of G and H*.

For our convenience, let us drop the notations $o, *$ and ., instead, we use only multiplication notation i.e.

For the groups G and H, we define, the group $G \times H$ where the operation is defined as

$$(g_1,h_1)(g_2,h_2) = (g_1g_2, h_1h_2)$$

as external direct product of G and H.

We can extent this concept to a finite number of groups, that is, if $G_1, G_2, \dots G_n$ are groups then we can make $G_1 \times G_2 \times \dots \times G_n$ into a group by means of a binary operation of multiplication by components. In other words, if $(a_1, a_2, \dots, a_n), (b_1, b_2, \dots, b_n) \in G_1 \times G_2 \times \dots \times G_n$ and we define

$$(a_1, a_2, \dots, a_3)(b_1, b_2, \dots, b_3) = (a_1b_1, a_2b_2, \dots, a_nb_n)$$

Then $G_1 \times G_2 \times \dots \times G_n$ is a group.

1.3.1 Example.

(a) Let us consider the group $(\mathbb{Z} \times \mathbb{Z}, +)$ where $'+'$ is defined component wise, that is, if $a = (x_1, y_1) \in \mathbb{Z} \times \mathbb{Z}$ and $b = (x_2, y_2) \in \mathbb{Z} \times \mathbb{Z}$ then

$a + b = (x_1 + x_2,\ y_1 + y_2)$. It is easy to verify that $(\mathbb{Z} \times \mathbb{Z}, +)$ is a group with $(0,0)$ as identity element.

(b) If we consider $\mathbb{Z}$, the additive group of integers, and $\mathbb{C}^*$, the multiplicative group of all non-zero complex numbers, then $G = \mathbb{Z} \times \mathbb{C}^*$ forms a group with respect to the binary composition defined as $x = (m_1, z_1), y = (m_2, z_2) \in G$ implies $xy = (m_1 + m_2, z_1z_2)$ where $(0,1)$ acts as the identity element of G and inverse of $(n,z) \in G$ will be $(-n, z^{-1})$.

(c) Consider the groups $(\mathbb{Z}_2, +)$ and $(\mathbb{Z}_3, +)$. Then $\mathbb{Z}_2 \times \mathbb{Z}_3$ forms an additive group where addition is defined component wise. Here,

$$\mathbb{Z}_2 \times \mathbb{Z}_3 = \{(0,0), (0,1), (0,2), (1,0), (1,1), (1,2)\}$$

Is it a cyclic group? Is it isomorphic to $\mathbb{Z}_6$? Consider the element $(1,1)$. Now,

$$\begin{aligned} 2(1,1) &= (1,1) + (1,1) = (0,2) \\ 3(1,1) &= 2(1,1) + (1,1) = (0,2) + (1,1) = (1,0) \\ 4(1,1) &= 3(1,1) + (1,1) = (1,0) + (1,1) = (0,1) \\ 5(1,1) &= 4(1,1) + (1,1) = (0,1) + (1,1) = (1,2) \\ 6(1,1) &= 5(1,1) + (1,1) = (1,2) + (1,1) = (0,0) \end{aligned}$$

Here, we see that, order of $(1,1)$ is 6, same as order of the group. Hence, $\mathbb{Z}_2 \times \mathbb{Z}_3$ is a cyclic group and is generated by $(1,1)$. Thus, the group $\mathbb{Z}_2 \times \mathbb{Z}_3$ is isomorphic to $\mathbb{Z}_6$ as all cyclic groups of order 6 are isomorphic to $\mathbb{Z}_6$.

(d) What happens for the group $\mathbb{Z}_2 \times \mathbb{Z}_2$ where the addition is defined component wise [here $(\mathbb{Z}_2, +)$ is a group] ?

$$\mathbb{Z}_2 \times \mathbb{Z}_2 = \{(0,0), (0,1), (1,0), (1,1)\}$$

Now,

$$\begin{aligned} 2(0,1) &= (0,1) + (0,1) = (0,0) \\ 2(1,0) &= (1,0) + (1,0) = (0,0) \\ 2(1,1) &= (1,1) + (1,1) = (0,0) \end{aligned}$$

Thus, we see that order of each element of $\mathbb{Z}_2 \times \mathbb{Z}_2$ other than identity is 2, not equal to the order of the group. Hence, $(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$ is not cyclic. Therefore, it is not isomorphic to $(\mathbb{Z}_4, +)$.

Is it isomorphic to $V = \{e, a, b, ab\}$, the Klein's 4 group? Yes, it is. You can define the mapping from V to $\mathbb{Z}_2 \times \mathbb{Z}_2$ by $e \to (0,0)$, $a \to (1,0)$, $b \to (0,1)$, $ab \to (1,1)$ and check it.

Please, note that, example (c) and (d) show that in some cases $\mathbb{Z}_m \times \mathbb{Z}_n$ is isomorphic to $\mathbb{Z}_{mn}$ and in some cases it is not.

1.4 PROPERTIES OF EXTERNAL DIRECT PRODUCTS

If the order of each element of a finite number of groups known, can we find the order of any element of the direct product of those groups? Following theorem will clarify it.

1.4.1 Theorem. The order of an element in a direct product of a finite number of groups is the least common multiple of the orders of the components of the element. In symbol,

$$o(g_1, g_2, \dots, g_n) = lcm\ (o(g_1), o(g_2), \dots, o(g_n))$$

Proof. Let $G = G_1 \times G_2 \times \dots \times G_n$ and e_i be the identity of G_i for $i = 1,2,\dots,n$.

Let $g = (g_1, g_2, \dots, g_n) \in G$.

Let $s = lcm\ (o(g_1), o(g_2), \dots, o(g_n))$ and $t = o(g)$. Now, s is a multiple of each $o(g_i)$ and hence,

$$g^s = (g_1, g_2, \dots, g_n)^s = (g_1^s, g_2^s, \dots, g_n^s) = (e_1, e_2, \dots, e_n) = e \in G$$

Therefore, $t \leq s$. Again.

$$(g_1^t, g_2^t, \dots, g_n^t) = (g_1, g_2, \dots, g_n)^t = e = (e_1, e_2, \dots, e_n)$$

So, $g_i^t = e_i$ for $i = 1,2,\dots,n$. Therefore, t is a common multiple of $o(g_i)$ for $i = 1,2,..,n$ but s is the least common multiple of $o(g_1), o(g_2),,, o(g_n)$. Therefore, $s \leq t$.

Hence, $s = t$ ∎

Look, theorem 1.4.1 is very helpful to solve the following problem.

1.4.2 How many elements of the group $\mathbb{Z}_{25} \times \mathbb{Z}_5$ are of order ?

Solution. Let $(a, b) \in \mathbb{Z}_{25} \times \mathbb{Z}_5$. Now, if $o(a, b) = 5$ then, we have,

$$o(a, b) = 5 = lcm\left(o(a), o(b)\right)$$

Therefore, either $o(a) = 5$ and $o(b) = 1$ or 5 or $o(a) = 1$ and $o(b) = 5$. We discuss two cases.

Case – 1.

$o(a) = 5$ and $o(b) = 1$ or 5.

There are only four elements in $\mathbb{Z}_{25}$ having order 5, namely, $\bar{5}$, $\overline{10}$, $\overline{15}$, $\overline{20}$.

If $o(b) = 1$ then $b = \bar{1}$. If $o(b) = 5$ then there are four possibilities, namely, $\bar{1}, \bar{2}, \bar{3}, \bar{4}$. Thus, there are only five choices for b. Therefore, number of elements in $\mathbb{Z}_{25} \times \mathbb{Z}_5$ is $4 \times 5 = 20$.

Case—2.

$o(a) = 1$ and $o(b) = 5$

Here, $a = \bar{1}$ and $b \in \{\bar{1}, \bar{2}, \bar{3}, \bar{4}\}$.

In this case, the number of elements in $\mathbb{Z}_{25} \times \mathbb{Z}_5$ with order 5 is $1 \times 4 = 4$.

Hence, there are only $20 + 4 = 24$ elements in $\mathbb{Z}_{25} \times \mathbb{Z}_5$ with order 5.

If G and H are two finite cyclic groups then is $G \times H$ cyclic? In general, the answer is negative as shown in Example (d) of 1.3. I think the following theorem will be helpful.

Theorem 1.4.3 Let G and H be two finite cyclic groups. Then $G \times H$ is cyclic if and only if $o(G)$ and $o(H)$ are prime to each other.

Proof. Let G and H be two cyclic groups such that $o(G) = m$ and $o(H) = n$. Then $o(G \times H) = mn$. Let $G = \langle g \rangle$ and $H = \langle h \rangle$. Then $o(g) = m$, $o(h) = n$.

Let us first suppose that $G \times H$ is cyclic and (g, h) be its generator. We shall show that $\gcd(m, n) = 1$. Let $\gcd(m, n) = d$. Now,

$$(g, h)^{\frac{mn}{d}} = \left((g^m)^{\frac{n}{d}}, (h^n)^{\frac{m}{d}}\right) = (e_G, e_H)$$

Since, $o(g, h) = mn$, we have, $mn \leq \frac{mn}{d}$ which shows that $d = 1$. Hence, $o(G)$ and $o(H)$ are prime to each other.

On the other hand, let $G = \langle g \rangle, H = \langle h \rangle$ and $\gcd(m, n) = 1$. Then,

$$o(g, h) = lcm\,(m, n) = mn = o(G \times H)$$

Thus $G \times H$ is cyclic and (g, h) being its generator.

Note. An external direct product $G_1 \times G_2 \times ... \times G_n$ of a finite number of finite cyclic groups is cyclic if and only if $o(G_i)$ and $o(G_j)$ are relatively prime for $i \neq j$.

1.5 THE GROUP OF UNITS MODULO n AS AN EXTERNAL DIRECT PRODUCT

If for $n > 1, U(n)$ be the set of all positive integers less than n and prime to n then $U(n)$ is a group under multiplication modulo n. This is known to us. Time has come for introduction of some new notations. If k divides n, let

$$U_k(n) = \{x \in U(n) : x \equiv 1(mod\ k)\}$$

For example, if you want $U_2(10)$, then first find $U(10)$ as $\{1,3,7,9\}$, then $U_2(10) = \{1,3,7,9\}$ but we see that $U_5(10) = \{1\}$. For another example, we have,

$U(21) = \{1, 2, 4, 5, 8, 10, 11, 13, 16, 17, 19, 20\}$, therefore,

$U_3(21) = \{1, 4, 10, 13, 16, 19\}$ and $U_7(21) = \{1, 8\}$.

You may ask whether $U_k(n)$ is a subgroup of $U(n)$. Yes, it is indeed because $1 \in U_k(n)$ and for $x, y \in U_k(n)$ we have, $x \equiv 1(mod\ k)$, $y \equiv 1(mod\ k)$ which shows that $xy \equiv 1(mod\ k)$. Since $U(n)$ is finite, we see that $U_k(n)$ is a subgroup of $U(n)$.

Theorem 1.5.1 For a given integer $n\ (> 1)$ let $n = st$ where $s, t \in \mathbb{Z}^+$ and $gcd\ (s, t) = 1$ then $U(n)$ or $U(st)$ is isomorphic to the external direct product of $U(s)$ and $U(t)$ i.e.

$$U(n) = U(st) = U(s) \times U(t)$$

Proof. We are given $\gcd(s, t) = 1$.

If $x \in U(st)$, then, $gcd(x, st) = 1 \Rightarrow gcd(x, s) = 1$ and $gcd(x, t) = 1$.

Let us define $\phi : U(st) \to U(s) \times U(t)$ by $\phi(x) = (x\ mod\ s, x\ mod\ t)$

We first show that ϕ is well defined. For that, let $x, y \in U(st)$ with $x = y$. Then $st|x - y$ which in turn implies $s|x - y, t|x - y$ as $\gcd(s, t) = 1$.

In other words, $x = y\ mod\ s, x = y\ mod\ t$. Thus,

$$x = y \Rightarrow (x\ mod\ s, x\ mod\ t) = (y\ mod\ s, y\ mod\ t) \Rightarrow \phi(x) = \phi(y)$$

Next we show that ϕ is one-one. Let $x, y \in U(st)$ such that $\phi(x) = \phi(y)$. Then

$(x\ mod\ s, x\ mod\ t) = (y\ mod\ s, y\ mod\ t) \Rightarrow x = y\ mod\ s$ and $x = y\ mod\ t$

So, $s|x - y$, $t|x - y$. Since, $\gcd(s, t) = 1$, we have, $st|x - y$.

Thus, $x \equiv y(mod\ st)$, i.e. $x = y$.

Now, we show that ϕ is onto. Let $(a, b) \in U(s) \times U(t)$. Then $\gcd(a, s) = 1$, $\gcd(b, t) = 1$.

Since, $\gcd(s,t) = 1$, there exists $m, n \in \mathbb{Z}$ such that $ms + nt = 1$ which shows that $\gcd(t, m) = 1$ and $\gcd(s, n) = 1$. Now, consider , $z = bsm + atn$.

We first show that $z \in U(st)$ i.e. $\gcd(z, st) = 1$.

Since st is composite, it must have a prime divisor p. Again, $p|st \Rightarrow p|s$ or $p|t$ as p is prime.

If $p|s$ then $p|bsm$ but p cannot divide atn, for if $p|atn$ then p must divide at least one of a, t or n but this is not the case as $\gcd(a, s) = \gcd(s, t) = \gcd(s, n) = 1$. So, p cannot divide z. Similar thing happens if $p|t$.

Hence, $\gcd(z, st) = 1$, in other words, $z \in U(st)$.

Now,

$z - a = bsm + atn - a = bsm - a(1 - tn) = bsm - ams = s(bm - am)$

So, $s|z - a$. Hence, $z = a \bmod s$. In the like manner, we have, $z = b \bmod t$.

Hence, $\phi(z) = (a \bmod s, b \bmod t)$.

Therefore, ϕ is onto.

Next, we shall show that ϕ is a homomorphism. Let $x, y \in U(st)$.

Thus, $\gcd(x, st) = 1 = \gcd(y, st)$. Hence, we have,

$$\gcd(x, s) = \gcd(x, t) = \gcd(y, s) = \gcd(y, t) = 1.$$

Then

$$\begin{aligned}\phi(xy) &= (xy \bmod s, xy \bmod t) = (x \bmod s, x \bmod t)(y \bmod s, y \bmod t)\\ &= \phi(x)\phi(y)\end{aligned}$$

Hence, ϕ is an isomorphism and we have, $U(st) \approx U(s) \times U(t)$ when $\gcd(s, t) = 1$.

Corollary : *If $m = n_1 n_2 \dots . n_k$ where $\gcd(n_i, n_j) = 1$ for $i \neq j$, then*

$$U(m) = U(n_1) \times U(n_2) \times \dots \times U(n_k).$$

How can we apply theorem 1.5.1? for example, we have, $70 = 2.5.7$ where 2,5,7 are prime to each other. Thus, by theorem 1.5.1,

$$\begin{aligned}U(70) &\approx U(2) \times U(5) \times U(7) \approx U(2) \times U(35) \approx U(10) \times U(7)\\ &\approx U(5) \times U(14)\end{aligned}$$

The order of any of the factors in the above can be interchanged.

We know that for each positive integer n, there is a cyclic group $\mathbb{Z}_n$ of order n. In 1801, Carl Gauss proved an important result :

$$U(2) \approx \{0\}, \quad U(4) \approx \mathbb{Z}_2, \quad U(2^n) \approx \mathbb{Z}_2 \times \mathbb{Z}_{2^{n-2}} \text{ for } n \geq 3$$

and $$U(p^n) \approx \mathbb{Z}_{p^n - p^{n-1}} \text{ for } p \text{ an odd prime.}$$

Keeping these results in mind and applying corollary to theorem 1.5.1, we can express any U –group as an external product of cyclic groups. For example,

$$U(70) \approx U(2) \times U(5) \times U(7) \approx \{0\} \times \mathbb{Z}_4 \times \mathbb{Z}_6$$

And $U(80) = U(16.5) \approx U(2^4) \times U(5) \approx \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_4$ [$U(2^4) \approx \mathbb{Z}_2 \times \mathbb{Z}_{2^{4-2}}$]

1.6 INTERNAL DIRECT PRODUCT

External direct product of groups gives a method by which we get a larger group from a number of groups so that we can determine some properties of the larger group from the properties of smaller groups. For example, if $G = H \times K$ then $o(G) = o(H).o(K)$. Any element of G can be written as (h, k) where $h \in H$ and $k \in K$; if $o(h)$ and $o(k)$ are finite then we have, $o(h, k) = lcm\left(o(h), o(k)\right)$. If H and K are abelian then G is abelian. If H and K are cyclic and $\gcd\left(o(H), o(K)\right) = 1$, then G is cyclic. Now, we may ask whether we can reverse this process—that is, given a larger group G, can we break it down into a product of subgroups in such a way that it would be possible to determine many properties of G from properties of the component pieces. Let's come to the following definition.

1.5.1 Definition : A group G is called the **internal direct product of $N_1, N_2, \dots, N_n$** if

(i) $N_1, N_2, \dots, N_n$ are normal subgroups of G
(ii) $G = N_1 N_2 \dots N_n$
(iii) Given $g \in G$ then $g = m_1 m_2 \dots m_n$, $m_i \in N_i$ $(i = 1,2,..,n)$ in a unique way.

Condition (iii) of the definition 1.5.1 explains that if $g \in G$ and

$$g = x_1 x_2 \dots x_n = y_1 y_2 \dots y_n$$

where $x_i, y_i \in N_i$ for $i = 1,2,..,n$ then $x_i = y_i$ for all $i = 1,2,\dots,n$.

1.5.2 Example.

1. Consider the Klein's four group $V = \{e, a, b, ab\}$, where $a^2 = b^2 = (ab)^2 = e$.

   Let $= \{e, a\}$, $K = \{e, b\}$. Clearly, H, K are subgroups of V and they are normal as V is abelian.

   Now, $H \cap K = \{e\}$ and $HK = \{e, ae, eb, ab\} = V$.

   Thus V is internal direct product of H and K.

2. Let $G = (\mathbb{Z}_6, +)$ and let $H = \{0,2,4\}$, $K = \{0,3\}$. (we omit the bar symbol)

   Then, clearly, H and K are normal subgroups of $\mathbb{Z}_6$. Now, we see that,

$$0 = 0 + 0, \quad 1 = 4 + 3, \quad 2 = 2 + 0$$
$$3 = 0 + 3 \quad 4 = 4 + 0 \quad 5 = 2 + 3$$

Therefore, $G = H + K$ and every element of G be expressed as $h + k$ where $h \in H, k \in K$ in a unique way.

Here is an important discussion. Suppose that G is the internal direct product of the normal subgroups $N_1, N_2, \dots, N_n$. We may consider $N_1, N_2, \dots, N_n$ as groups (let us forget that they are normal subgroups). As per our previous knowledge, we can form the external direct product of $N_1, N_2, \dots, N_n$ as

$$T = N_1 \times N_2 \times \dots \times N_n$$

Is there any relation between G and T? Yes, my dear reader, there is a relation. We shall show that G is isomorphic to T and if it is established then we can omit the prefixes *internal*and *external,* because upto isomorphism there would be no difference between external direct product and internal direct product of groups.

1.5.2 Lemma : If G is the internal direct product of $N_1, \dots, N_2$ and if $a \in N_i, b \in N_j$ for $i \neq j$ then $N_i \cap N_j = \{e\}$and $ab = ba$.

Proof. Let $x \in N_i \cap N_j$. Then $x \in G$ and we can write

$$x = e_1 e_2 \dots e_{i-1} x e_{i+1} \dots e_{j-1}\ e_j\ e_{j+1} \dots e_n$$

where $e_k = e$ being treated as the identity of N_k for $k = 1,2, \dots, n$ and $x \in N_i$.

Again, we can write,

$$x = e_1 \dots e_{i-1} e_i e_{i+1} \dots e_{j-1}\ x\ e_{j+1} \dots e_n$$

Since, expression of x is unique, we have $x = e_i = e$.

Hence, $N_i \cap N_j = \{e\}$.

Let $a \in N_i, b \in N_j$ where $i \neq j$. Since N_j is a normal subgroup of G, we have, $aba^{-1} \in N_j$.

Again, $b \in N_j \Rightarrow b^{-1} \in N_j$. Thus, $aba^{-1}b^{-1} \in N_j$, by closure property.

Similarly, $a \in N_i \Rightarrow a^{-1} \in N_i$. As N_i is a normal subgroup of G, we have, $ba^{-1}b^{-1} \in N_i$.

Therefore, $aba^{-1}b^{-1} \in N_i$. Hence, $aba^{-1}b^{-1} \in N_i \cap N_j = \{e\}$.

Thus, $aba^{-1}b^{-1} = e$ i.e. $ab(ba)^{-1} = e$ which shows that

$$ab = ba.$$

Now, let us try to prove the much awaited theorem.

1.5.3 Theorem. Let G be a group and G be the internal direct product of $N_1, N_2, \dots, N_n$. Let T be the external direct product of $N_1, N_2, \dots N_n$, that is, $T = N_1 \times N_2 \times \dots \times N_n$. Then G and T are isomorphic.

Proof. Let us define a mapping $f : T \to G$ by

$$f(x_1, x_2, \dots, x_n) = x_1 x_2 \dots x_n$$

where each $x_i \in N_i$ for $i = 1,2, \dots, n$. We shall show that f is an isomorphism of T onto G.

Let $a, b \in T$ where $a = (a_1, a_2, \dots, a_n)$, $b = (b_1, b_2, \dots, b_n)$. Now,

$$\begin{aligned} f(ab) &= f\big((a_1, a_2, \dots, a_n)(b_1, b_2, \dots, b_n)\big) \\ &= f(a_1 b_1, a_2 b_2, \dots, a_n b_n) \\ &= a_1 b_1 a_2 b_2 \dots a_n b_n \end{aligned}$$

But by Lemma 1.5.2, $a_i b_j = b_j a_i$ for $i \neq j$. Thus, we have,

$$a_1 b_1 a_2 b_2 \dots a_n b_n = a_1 a_2 \dots a_n b_1 b_2 \dots b_n.$$

Therefore,

$$f(ab) = a_1 b_1 a_2 b_2 \dots a_n b_n = a_1 a_2 \dots a_n b_1 b_2 \dots b_n = f(a) f(b)$$

So, f is a homomorphism.

Now,

$$f(a_1, a_2, \dots, a_n) = f(b_1, b_2, \dots, b_n) \Rightarrow a_1 a_2 \dots a_n = b_1 b_2 \dots b_n$$

By the uniqueness in the definition of internal product, we have,

$a_1 = b_1$, $a_2 = b_2, \dots, a_n = b_n$, in other words, $(a_1, a_2, \dots, a_n) = (b_1, b_2, \dots, b_n)$.

Therefore, f is one-one.

Now, let $x \in G$. Then $x = a_1 a_2 \dots a_n$ where $a_1 \in N_1, a_2 \in N_2, \dots, a_n \in N_n$. Therefore,

$$f(a_1, a_2, \dots, a_n) = a_1 a_2 \dots a_n = x$$

So, f is onto.

Hence, f is an isomorphism.

Therefore, G and T are isomorphic ■

1.6 CONVERSE OF LAGRANGE'S THEOREM FOR FINITE ABELIAN GROUPS

It has already been studied in my book *Group Theory 1* that if G is a finite group then order of any subgroup H of G divides the order of G(Lagrange's Theorem). What about the converse? That is, if order of a group G be n and d is a positive divisor of n, then can we have a subgroup of G having order d? Not sure. For example, order of A_4, the group of even permutations of a set containing 4

elements with respect to the composition of permutations, is 12. Now, 6 is a divisor of 12. Does there exist a subgroup H of A_4 containing 6 elements? If it does, let's see what happens? Let H be a subgroup of A_4 such that $o(H) = 6$. Then $[A_4 : H] = \frac{o(A_4)}{o(H)} = \frac{12}{6} = 2$. Therefore, H is a normal subgroup of A_4 and $A_4/H \cong \mathbb{Z}_2$. Since the order of the quotient group A_4/H is 2, the square of each element of the group A_4/H must be identity, that is, for all $x \in A_4$, $(xH)^2 = H$ or $x^2H = H$. That is for all $x \in A_4$, we have, $x^2 \in H$. Let g be an element of A_4 of order 3. Therefore, $g = (g^2)^2 [\text{ as } g^3 = e]$. Now, $g \in A_4 \Rightarrow g^2 \in A_4 \Rightarrow (g^2)^2 \in H$. Thus, $g \in H$. Therefore, it is shown that H must contain all elements of A_4 of order 3. This is a contradiction as there are 8 elements of A_4 which are of order 3 whereas order of H is 6 (< 8).

Hence, it is clear that converse of Lagrange's theorem is not true, in general. But Cauchy showed that the converse of Lagrange's theorem holds if we consider the group as finite abelians. Thus, if G is an abelian group of order n and d is a positive divisor of n, then G must have a subgroup of order d. But before that we wish to offer a very important theorem for finite abelian groups due to Cauchy.

1.6.1 Theorem (Cauchy). Let G be an abelian group of order n and p be a prime divisor of n. Then G has an element of order p or equivalently, G has a subgroup of order p.

Proof. We shall use induction on $n = o(G)$.

If $o(G) = p$, a prime, then $o(a) = p$ for all $a \in G - \{e\}$. Hence, the result is true if $o(G) = 2$.

Let us assume that the result holds for all abelian groups of order r, where $2 \leq r < n$. Let $o(G) = n$. By our assumption, if for some proper subgroup H of G, p divides $o(H)$, then H (and hence G) must have an element of order p.

Thus, we assume that if H is a proper subgroup of G then p does not divide $o(H)$. Since, H is a subgroup of an abelian group, H is a normal subgroup of G and hence we have a quotient group G/H. Now, we know that $o(G) = o(H).o(G/H)$.

Since p divides $o(G)$ and p does not divide $o(H)$, it is clear that p divides $o(G/H)$. As $o(G/H) < n$, by induction hypothesis, G/H has an element, say aH, of order p. That is, $(aH)^p = H$ and hence $a^p \in H$.

If $o(H) = m$ then $(a^p)^m = e$ i.e. $(a^m)^p = e$, i.e. $b^p = e$ where $b = a^m \in G$.

We claim that $b \neq e$. If $b = a^m = e$ then we have, $(aH)^m = a^m H = eH = H$.

Since, $\gcd(p, m) = 1$, there exist $u, v \in \mathbb{Z}$ such that $pu + mv = 1$. Thus,

$$aH = a^{pu+mv}H = (aH)^{pu}(aH)^{mv} = H \,[\text{ as } (aH)^p = H = (aH)^m]$$

which is a contradiction as $o(aH) = p$ and p does not divide $o(H)$.

Hence, $b = a^m \neq e$. Therefore, $b = a^m$ is an element of G whose order is p and $H = \{(a^m)^t : t \in \mathbb{Z}\}$ is a subgroup of G having order p ■

Note. Cauchy's theorem can be extended to any finite group (including non-abelian groups), that is, if order of a group be n and p be a prime divisor of n, then the group has a subgroup of order p. But it is beyond the scope of the syllabus.

Now, let us come to the original problem, that is, converse of Lagrange's theorem, which is not true, in general, but holds in case of finite abelian groups.

1.6.2 Theorem (Converse of Lagrange's theorem for finite abelian groups)
Let G be an abelian group of order n and m be a positive divisor of n, then G has a subgroup of order m.

Proof. If $m = 1$, then $\{e\}$ serves our purpose. If $n = m = 1$, then $G = \{e\}$ and the result is obvious. So, let $n > 1$, $m > 1$. We shall use induction on n.

If $n = 2$, then $m = 2$ and G itself serves the required subgroup of order m. Hence, the result is true for $n = 2$.

Let us assume that the result holds for all abelian groups of order r with $2 \leq r < n$.

Now, m must have some prime divisor, say p. So, there exists $q \in \mathbb{Z}$ such that $m = p.q$.

Clearly, p is a prime divisor of n. Hence, by Cauchy's theorem, G must have a subgroup H of order p.

Since, G is commutative, H is normal in G and therefore the quotient group G/H exists. Now,

$$1 \leq o(G/H) = \frac{o(G)}{o(H)} < o(G)$$

Here, $o(G/H)$ is $\frac{n}{p}$. Since, m divides n, there exists $t \in \mathbb{Z}$, such that $n = mt$.

Therefore,

$$o(G/H) = \frac{n}{p} = \frac{mt}{p} = \frac{pqt}{p} = qt$$

which shows that q divides $o(G/H)$. Hence, by induction hypothesis, G/H has a subgroup K/H such that $o(K/H) = q$, where K is a subgroup of G. Now,

$$o(K) = o(K/H).o(H) = qp = m$$

Therefore, G has a subgroup K such that $o(K) = m$.

Hence proved ■

1.7 FUNDAMENTAL THEOREM OF FINITE ABELIAN GROUPS

Now, we state a theorem, known as fundamental theorem of abelian groups, that describes all finite abelian groups in a standardized way.

1.7.1 Theorem : Every finite abelian group is a direct product of cyclic groups of prime-power order. Moreover, the number of terms in the product and the orders of the cyclic groups are uniquely determined by the group.

The proof is too long and difficult to offer and hence is omitted.

Since a cyclic group of order n is isomorphic to $\mathbb{Z}_n$, theorem 1.7.1 states that a finite abelian group G is isomorphic to a group of the form

$$\mathbb{Z}_{p_1^{n_1}} \times \mathbb{Z}_{p_2^{n_2}} \times \dots \times \mathbb{Z}_{p_k^{n_k}}$$

where the $p_i's$ are not necessarily distinct primes and the prime powers $p_1^{n_1}, p_2^{n_2}, \dots, p_k^{n_k}$ are uniquely determined by G.

Solved Problems :

1. **If A and B two groups, then prove that $A \times B$ is isomorphic to $B \times A$.**

Solution. Let us define a map $f : A \times B \to B \times A$ by

$$f(a,b) = (b,a)$$

Let $x, y \in A \times B$ where $x = (a,b)$ and $y = (c,d)$.

Then $xy = (ac, bd)$ and $f(x) = (b,a), f(y) = (d,c)$. Now,

$$f(xy) = f(ac,bd) = (bd,ac) = (b,a)(d,c) = f(x)f(y)$$

So, f is a homomorphism.

Again, $f(a,b) = f(c,d) \Rightarrow (b,a) = (d,c) \Rightarrow b = d,$

$$a = c \Rightarrow (a,b) = (c,d)$$

Thus, f is injective.

f is onto, because for any $(a,b) \in B \times A$, there exists $(b,a) \in A \times B$ such that $f(b,a) = (a,b)$.

Therefore, f is an isomorphism.

Hence, $A \times B \cong B \times A$.

(In general, the external direct product of any number of groups is isomorphic to the external direct product of any rearrangement of those groups.)

2. **Prove that the group of complex numbers under addition is isomorphic to $\mathbb{R} \times \mathbb{R}$.**

Solution. Let $\mathbb{C}$ denote the group of complex numbers under addition.

Define $f : \mathbb{C} \to \mathbb{R} \times \mathbb{R}$ by $f(a+ib) = (a,b)$.

Let $a+ib, c+id \in \mathbb{C}$. Then

$$\begin{aligned} f[(a+ib)+(c+id)] &= f[(a+c)+i(b+d)] \\ &= (a+c, b+d) \\ &= (a,b)+(c,d) \\ &= f(a+ib)+f(c+id) \end{aligned}$$

So, f is a homomorphism.

Let $a+ib, c+id \in \mathbb{C}$ such that $f(a+ib) = f(c+id)$, that is, $(a,b) = (c,d)$

Thus, $a = c, b = d$. Hence, $f(a+ib) = f(c+id) \Rightarrow a+ib = c+id$.

Therefore, f is injective.

For $(a,b) \in \mathbb{R} \times \mathbb{R}$, there exists $a+ib \in \mathbb{C}$ such that $f(a+ib) = (a,b)$.

So, f is onto.

Hence, f is an isomorphism. In other words, $\mathbb{C} \cong \mathbb{R} \times \mathbb{R}$.

3. **Prove or disprove : $\mathbb{C}^*$, the group of non-zero complex numbers under multiplication, is isomorphic to $\mathbb{R}^* \times \mathbb{R}^*$ where $\mathbb{R}^*$ denotes the group of non-zero real numbers under addition.**

Solution. The statement is wrong. Because, in $\mathbb{R}^* \times \mathbb{R}^*$ there are three elements of order 2, viz. $(1,-1), (-1,1), (-1,-1)$ whereas, $\mathbb{C}^*$ contains only one element of order 2, that is, -1. This cannot happen as isomorphism preserves order.

4. **Prove or disprove : $\mathbb{Z} \times \mathbb{Z}$ is a cyclic group.**

Solution. The statement is false. We know, that $\mathbb{Z} = \langle 1 \rangle$ but $(1,1)$ is not a generator of $\mathbb{Z} \times \mathbb{Z}$ as $(1,2) \notin \langle (1,1) \rangle$. Hence, $\mathbb{Z} \times \mathbb{Z}$ is not cyclic.

5. **Is $\mathbb{Z}_3 \times \mathbb{Z}_9$ is isomorphic to $\mathbb{Z}_{27}$?**

Solution. No, as $\gcd(3,9) = 3 \neq 1$. We know that $\mathbb{Z}_m \times \mathbb{Z}_n$ is isomorphic to $\mathbb{Z}_{mn}$ if and only if $\gcd(m,n) = 1$.

6. **For each integer $n > 1$, give examples of two non-isomorphic groups of order n^2.**

Solution. $\mathbb{Z}_{n^2}$ and $\mathbb{Z}_n \times \mathbb{Z}_n$ as $\mathbb{Z}_{n^2}$ is cyclic whereas $\mathbb{Z}_n \times \mathbb{Z}_n$ is not cyclic as $\gcd(n,n) = n > 1$.

7. **Let G be a group with identity e_G and H be a group with identity e_H. Prove that G is isomorphic to $G \times \{e_H\}$ and H is isomorphic to $\{e_G\} \times H$.**

Solution. Let us define a map $f : G \to G \times \{e_H\}$ by $f(g) = (g, e_H)$

If $g_1, g_2 \in G$, we have

$$f(g_1 g_2) = (g_1 g_2, e_H) = (g_1, e_H)(g_2, e_H) = f(g_1)f(g_2)$$

So, f is a homomorphism.

Let $g_1, g_2 \in G$ such that $f(g_1) = f(g_2)$.

Then $(g_1, e_H) = (g_2, e_H)$, that is, $g_1 = g_2$

So, $f(g_1) = f(g_2) \Rightarrow g_1 = g_2$, in other words, f is injective.

For any $(g, e_H) \in G \times \{e_H\}$, there exists $g \in G$ such that $f(g) = (g, e_H)$.

So, f is onto.

Thus, f is an isomorphism. Hence, $G \cong G \times \{e_H\}$.

Similarly, taking $\phi : H \to \{e_G\} \times H$ defined by $\phi(h) = (e_G, h)$, it can be shown that $H \cong \{e_G\} \times H$.

8. If $G \times H$ is cyclic, prove that G and H is cyclic.

Solution. By problem 7, it is clear that $G \cong G \times \{e_H\}$. Now, $G \times \{e_H\}$ is a subgroup of $G \times H$ and hence is cyclic as any subgroup of a cyclic group is cyclic. Thus, G is cyclic. Similarly, it can be proved that H is cyclic.

9. If a group has exactly 24 elements of order 6, how many cyclic subgroups of order 6 does it have?

Solution. Let G be a group having exactly 24 elements of order 6. Now, any cyclic group of order 6 is isomorphic to $(\mathbb{Z}_6, +)$. Since, $\mathbb{Z}_6$ has exactly two generators, any cyclic group of order 6 has exactly 2 generators. Thus G can have exactly 12 subgroups of order 6.

10. If an abelian group G is the internal direct product of its subgroups H and K, then prove that $H \cong G/K$ and $K \cong G/H$.

Solution. Let $g \in G$. Then $g = hk$ where $h \in H, k \in K$.

Let us define $f : G \to H$ by $f(g) = h$ and $F : G \to K$ by $F(g) = k$.

Let $g_1, g_2 \in G$. Then $g_1 = h_1 k_1$, $g_2 = h_2 k_2$ where, $h_1, h_2 \in H$ and $k_1, k_2 \in K$. Now,

$$f(g_1 g_2) = f(h_1 k_1 h_2 k_2) = f(h_1 h_2 k_1 k_2) \text{ [as } G \text{ is abelian]}$$
$$= h_1 h_2 = f(g_1)f(g_2).$$

Therefore, f is a homomorphism and f is onto by definition.

$$x \in ker f \Leftrightarrow f(x) = e \Leftrightarrow x = ek, \forall k \in K$$

Thus, $ker f = K$. Hence, by 1st isomorphism theorem, $G/ker f \cong H$, i.e. $G/K \cong H$.

Similarly, it can be proved that $G/H \cong K$.

11. If $T = G_1 \times G_2 \times \ldots \times G_n$ prove that for each $i = 1, 2, \ldots, n$ there is a homomorphism f_i of T onto G_i. Find the kernel of f_i.

Solution. For some $i \in \{1, 2, \ldots, n\}$, let us define $f_i : T \to G_i$ by

$$f_i(g_1, g_2, \ldots, g_{i-1}, g_i, g_{i+1} \ldots g_n) = g_i$$

Let $(g_1, g_2, \ldots, g_{i-1}, g_i, g_{i+1} \ldots g_n), (h_1, h_2, \ldots, h_{i-1}, h_i, h_{i+1} \ldots h_n) \in T$. Then

$f_i(g_1, g_2, \ldots, g_{i-1}, g_i, g_{i+1} \ldots g_n) = g_i$, $\quad f_i(h_1, h_2, \ldots, h_{i-1}, h_i, h_{i+1}, \ldots h_n) = h_i$

Now,

$$f_i[(g_1, g_2, \ldots, g_{i-1}\, g_i\, g_{i+1} \ldots g_n)(h_1, h_2, \ldots, h_{i-1}\, h_i\, h_{i+1} \ldots h_n)]$$
$$= f_i(g_1 h_1, \ldots, g_{i-1} h_{i-1}, g_i h_i, g_{i+1} h_{i+1}, \ldots, g_n h_n) = g_i h_i$$
$$= f_i(g_1, g_2, \ldots, g_{i-1}, g_i, g_{i+1} \ldots g_n) f_i(h_1, h_2, \ldots, h_{i-1}, h_i, h_{i+1}, \ldots h_n)$$

which shows that f_i is a homomorphism.

For any $g_i \in G_i$, there exists $(g_1, g_2, \ldots, g_{i-1}, g_i, g_{i+1} \ldots g_n) \in T$ such that

$$f_i(g_1, g_2, \ldots, g_{i-1}, g_i, g_{i+1} \ldots g_n) = g_i$$

Hence, f_i is onto.

Clearly, $ker f_i$ is given by

$$ker f_i = \{(g_1, g_2, \ldots, g_{i-1}, e_i, g_{i+1} \ldots g_n) : g_k \in G_k, k = 1, 2, \ldots, n\}$$

where e_i is the identity element of G_i.

12. What is the order of any non-identity element of $\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3$?

Solution. We know that order of any non-identity element of $\mathbb{Z}_3$ is 3 (in fact, order of any non-identity element of $\mathbb{Z}_p$ is p, p being prime). Let $(a, b, c) \in \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3$ where $(a, b, c) \neq (e, e, e)$.

Now, order of (a, b, c) is $lcm\{o(a), o(b), o(c)\} = 3$.

13. What is the largest order of any element in $\mathbb{Z}_{30} \times \mathbb{Z}_{20}$?

Solution. Order of $(1,1) = lcm\,\{30, 20\} = 60$.

14. Let G_1 and G_2 be two cyclic groups of order 2 and 3 respectively. Prove that $G = G_1 \times G_2$ is a cyclic group of order 6.

Solution. Let $G_1 = \langle a \rangle$ and $G_2 = \langle b \rangle$. Since order of G_1 and G_2 are 2 and 3 respectively, we have, $o(a) = 2$, $o(b) = 3$ and we can write

$$G_1 = \{e_1, a\}, G_2 = \{e_2, b, b^2\}$$

where e_1 and e_2 are the identity elements of G_1 and G_2 respectively. Now,

$$G = G_1 \times G_2 = \{(e_1, e_2), (e_1, b), (e_1, b^2), (a, e_2), (a, b), (a, b^2)\}$$

It is clear that G is a group with respect to the operation $(a,b)(c,d) = (ac,bd)$.

Let $g = (a,b) \in G$. Then

$$g^2 = (a^2,b^2) = (e_1,b^2),\ g^3 = (a^3,b^3) = (a,e_2),\ g^4 = (a^4,b^4) = (e_1,b)$$
$$g^5 = (a^5,b^5) = (a,b^2),\ g^6 = (a^6,b^6) = (e_1,e_2)$$

Hence, $G = < g >$ is a cyclic group of order 6.

15. Let G be a group and $T = G \times G$. Then show that
(a) $D = \{(g,g) \in G \times G : g \in G\}$ is isomorphic to G and
(b) D is normal in T if and only if G is abelian.

Solution.

(a) It is easy to show that D is a subgroup of T.

Let us define a map $f : D \to G$ by $f(g,g) = g$. Now,

$$f[(g,g)(h,h)] = f(gh,gh) = gh = f(g,g)f(h,h)$$

So, f is a homomorphism.

To show f is injective,let $(g,g),(h,h) \in D$ such that $f(g,g) = f(h,h)$. Then $g = h$ which implies $(g,g) = (h,h)$, that is, f is one-one.

By definition, f is onto.

Hence, f is an isomorphism. In other words, $D \cong G$.

(b) Let D be a normal subgroup of T.Let $a,b \in G$. Then $(a,e) \in T$ and $(b,b) \in D$, e being the identity element of G. Since, D is normal in T we have,

$$(a,e)(b,b)[(a,e)]^{-1} \in D => (aba^{-1}, ebe^{-1}) \in D$$
$$=> aba^{-1} = b => ab = ba$$

Therefore, G is abelian.

Conversely, let G be abelian. Let $(a,b) \in T$ and $(g,g) \in D$

Now, $$(a,b)(g,g)[(a,b)]^{-1} = (a,b)(g,g)(a^{-1},b^{-1})$$
$$= (aga^{-1}, bgb^{-1} = (g,g) \in D$$
$$[\textit{as } G \textit{ is abelian}, aga^{-1} = aa^{-1}g = g, bgb^{-1} = g]$$

Hence, D is a normal subgroup of T.

16. Give an example of a group G and normal subgroups $N_1, N_2, \ldots, N_n$ such that$G = N_1N_2 \ldots N_n$ and $N_i \cap N_j = \{e\}$ for $i \neq j$ and yet G is *not* the internal direct product of $N_1, N_2, \ldots, N_n$.

olution. Let $G = \{e, a, a^2, b, b^2, ab, a^2b^2\}$ where $ab = ba$, $a^3 = b^3 = e$. Then G an abelian group. Let $N_1 = \{e, a, a^2\}$, $N_2 = \{e, ab, a^2b^2\}$, $N_3 = \{e, b, b^2\}$. Then $_1, N_2, N_3$ are subgroups of G and each of them is normal as G is abelian.

Clearly, $G = N_1N_2N_3$ and $N_i \cap N_j = \{e\}$ for $i \neq j$. But we see that $ab \in G$ and $ab = eabe = aeb$

That is, ab has two different representation. Hence, G is not the internal direct product of N_1, N_2, N_3 as in that case each element of G would have unique representation, which is not the case.

17. Let G be a group and $K_1, K_2, \ldots, K_n$ normal subgroups of G. Suppose that $K_1 \cap K_2 \cap \ldots \cap K_n = \{e\}$. Let $V_i = G/K_i$ for $i = 1, 2, \ldots, n$. Prove that there is an isomorphism of G into $V_1 \times V_2 \times \ldots \times V_n$.

Solution. Let us define a map $f : G \to V_1 \times V_2 \times \ldots \times V_n$ by

$$f(g) = (gK_1, gK_2, \ldots, gK_n)$$

For $g, h \in G$, we have,

$$f(gh) = (ghK_1, ghK_2, \ldots, ghK_n)$$
$$= (gK_1, gK_2, \ldots, gK_n)(hK_1, hK_2, \ldots, hK_n)$$
$$= f(g)f(h)$$

Therefore, f is a homomorphism.

To show f is injective, let $g, h \in G$ such that $f(g) = f(h)$, that is,

$$(gK_1, gK_2, \ldots, gK_n) = (hK_1, hK_2, \ldots, hK_n)$$

So, $gK_1 = hK_1, gK_2 = hK_2, \ldots, gK_n = hK_n$. Thus,

$$g^{-1}h \in K_1, g^{-1}h \in K_2, \ldots, g^{-1}h \in K_n$$

Whichshows that $g^{-1}h \in K_1 \cap K_2 \cap \ldots \cap K_n = \{e\}$(given).

Hence, $g^{-1}h = e, i.e.\ g = h$.

Therefore, f is one-one.

Hence, G is isomorphic to $f(G)$, a subgroup of $V_1 \times V_2 \times \ldots \times V_n$.

18. Show that every group of order p^2, p a prime, is either cyclic or is isomorphic to the direct product of two cyclic groups each of order p.

Solution. We know that every group of order p^2 is abelian. Let G be a group of order p^2.

Order of each element of G is either 1 or p or p^2.

If there is an element in G of order p^2, then G is cyclic.

So, let there be no elements in G of order p^2. Let $h \in G, h \neq e$. Then $o(h) = p$.

Let $H = \{h^n : n \in \mathbb{Z}\}$. So, $o(H) = p$.

Thus, $G - H \neq \emptyset$. Let $k \in G - H$ and $K = \{k^n : n \in \mathbb{Z}\}$ and hence $o(K) = p$.

Now, H and K are two subgroups of G and they are normal as G is abelian.

Clearly, $o(G) = o(H)o(K)$ and $H \cap K = \{e\}$. So, $G = HK$.

Let $x \in G$. If x has two expressions like $x = h^q k^r = h^s k^t$,
then $h^{q-s} = k^{t-r} = e$ as $H \cap K = \{e\}$.

So, each element of G has unique representation. Therefore, G is the interna direct product of H and K. Hence, we can say that, G is isomorphic to the direct product of two cyclic groups each of order p.

19. If $G = K_1 \times K_2 \times ... \times K_n$ describe the centre of G in terms of those of the K_i.

Solution. Let $Z(K_i)$ denote the centre of K_i. Thus,

$$k_i \in Z(K_i) => k_i g_i = g_i k_i, \qquad \forall g_i \in K_i$$

We claim that centre of G, denoted by $Z(G)$, is given by

$$Z(G) = Z(K_1) \times Z(K_2) \times ... \times Z(K_n)$$

as $(k_1, k_2, .., k_n) \in Z(K_1) \times Z(K_2) \times ... \times Z(K_n)$ implies for any $(g_1, g_2, .., g_n) \in G$

$$\begin{aligned}(k_1, k_2, ... k_n)(g_1, g_2, ..., g_n) &= (k_1 g_1, k_2 g_2, ..., k_n g_n) \\ &= (g_1 k_1, g_2 k_2, ..., g_n k_n) \\ &= (g_1, g_2, .., g_n)(k_1, k_2, .., k_n).\end{aligned}$$

20. Suppose that ϕ is an isomorphism from $\mathbb{Z}_3 \times \mathbb{Z}_5$ to $\mathbb{Z}_{15}$ and $\phi(2,3) = 2$. Find the element in $\mathbb{Z}_3 \times \mathbb{Z}_5$ that maps to 1.

Solution. Since, ϕ is an isomorphism and $\phi(2,3) = 2$, we have,

$$8\phi(2,3) = 16 = 1 \ (in\ \mathbb{Z}_{15})$$

Now, $8\,\phi(2,3) = \phi(16,24) = \phi\,(1,4)$. The required element in $\mathbb{Z}_3 \times \mathbb{Z}_5$ is (1,4) ■

21. What is the largest order of any element in $U(900)$?

Solution. We know that

$$U(2) \approx \{0\}, \quad U(4) \approx \mathbb{Z}_2, \quad U(2^n) \approx \mathbb{Z}_2 \times \mathbb{Z}_{2^{n-2}} \ for\ n \geq 3$$

and $U(p^n) \approx \mathbb{Z}_{p^n - p^{n-1}}$ for p an odd prime.

Now, $900 = 4 \times 3^2 \times 5^2$.

Thus, $U(900) = U(4 \times 3^2 \times 5^2) \approx U(4) \times U(3^2) \times U(5^2) \approx \mathbb{Z}_2 \times \mathbb{Z}_6 \times \mathbb{Z}_{20}$.

Hence, largest order of any element is $lcm\ \{2,6,20\} = 60$.

22. Give an example to show that there exists a group with elements a and b such that $o(a) = \infty, o(b) = \infty$ but $o(ab) = 2$.

Solution. Let us consider the group $\mathbb{Z} \times \mathbb{Z}_2$. Take $(1,1), (-1,0) \in \mathbb{Z} \times \mathbb{Z}_2$.

Then $o(1,1) = \infty$, $o(-1,0) = \infty$ but

$$o((1,1)(-1,0)) = o(0,1) = 2$$

as $(0,1) + (0,1) = (0,0)$ in $\mathbb{Z} \times \mathbb{Z}_2$.

23. Let p, q be odd primes and let m and n be positive integers, then check whether $U(p^m) \times U(q^n)$ is cyclic.

Solution. Since, p, q are odd primes and $m, n \in \mathbb{N}$, we have,

$$U(p^m) \approx \mathbb{Z}_{p^m - p^{m-1}} \approx \mathbb{Z}_{p^{m-1}(p-1)}$$

And $U(q^n) \approx \mathbb{Z}_{q^{n-1}(q-1)}$

Thus, $U(p^m) \times U(q^n) \approx \mathbb{Z}_{p^{m-1}(p-1)} \times \mathbb{Z}_{q^{n-1}(q-1)}$

which shows that $U(p^m) \times U(q^n)$ is not cyclic as $\mathbb{Z}_{p^{m-1}(p-1)} \times \mathbb{Z}_{q^{n-1}(q-1)}$ is not cyclic as

$\gcd(p^{m-1}(p-1), q^{n-1}(q-1)) \geq 2$ (as $p-1, q-1$ both are even) [Since we know that $\mathbb{Z}_m \times \mathbb{Z}_n$ is cyclic if and only if $\gcd(m,n) = 1$]

24. Check whether $U(55) \approx U(75)$.

Solution. We have, $U(55) = U(5.11) \approx U(5) \times U(11) \approx \mathbb{Z}_4 \times \mathbb{Z}_{10}$ [$U(p) \approx \mathbb{Z}_{p-1}$]

again, $U(75) = U(5^2.3) \approx U(5^2) \times U(3) \approx \mathbb{Z}_{(5^2-5)} \times \mathbb{Z}_2$

[$U(p^n) = \mathbb{Z}_{p^n - p^{n-1}}$] $\approx \mathbb{Z}_{20} \times \mathbb{Z}_2 \approx \mathbb{Z}_4 \times \mathbb{Z}_5 \times \mathbb{Z}_2 \approx \mathbb{Z}_4 \times \mathbb{Z}_{10}$

Hence, $U(55) \approx U(75)$.

25. What is the smallest positive integer k for which $x^k = e$ for all x in $U(pq)$ where p, q are distinct primes.

Solution. We have, $U(pq) \approx U(p) \times U(q) \approx \mathbb{Z}_{p-1} \times \mathbb{Z}_{q-1}$

Thus, $k = lcm\,\{p-1, q-1\}$.

26. Let p be a prime. Show that if H is a subgroup of a group of order $2p$ that is not normal, then H has order 2.

Solution. Let G be a group of order $2p$. Clearly, G and $\{e\}$ are normal. Using Lagrange's theorem we can say that the order of all other subgroups have either p or 2. If H is a subgroup of order p, then we have,

$$[G : H] = \frac{o(G)}{o(H)} = \frac{2p}{p} = 2$$

and hence, H becomes normal in G.
Therefore, any subgroup of G that is not normal in G, must be of order 2.

27. **Let H, K be subgroups of a commutative group G. Let $o(H) = s$ and $o(K) = t$. Let $d = lcm\{s, t\}$. Show that G has a subgroup of order d.**

Solution. Given G is commutative. Therefore, $HK = KH$ and hence, HK is a subgroup of G. Now, order of HK is finite as $o(H)$ and $o(K)$ are finite.

Since, H and K are subgroups of HK, by Lagrange's theorem, we have, $m|o(HK)$, $n|o(HK)$.

So, $d|o(HK)$. Since, HK is a finite commutative group and d is a divisor of $o(HK)$, we assert that HK has a subgroup of order d.

Hence, G has a subgroup of order d.

28. **Let G be a finite commutative group and k be a positive divisor of $o(G)$. Let**

$H = \{x \in G : x^k = e\}$. Prove that $o(H)$ is a multiple of k.

Solution. Since G is a finite commutative group and k is a positive divisor of $o(G)$, by converse of Lagrange's theorem for commutative groups, G has a subgroup of order k. Let it be K, that is, K is a subgroup of G such that $o(K) = k$

Now, $x \in K \Rightarrow x^k = e \Rightarrow x \in H$. Thus, $K \subseteq H$. Hence, by Lagrange's theorem, $o(K)|o(H)$, i.e., $k|o(H)$. In other words, $o(H) = kt$ for some positive integer t.

Hence, $o(H)$ is a multiple of k.

29. **Let G be an abelian group and $a, b \in G$ be of order m and n respectively where $gcd\ (m, n) = 1$. Show that there exists an element c of G such that $o(c) = k$ where k is the LCM of m and n.**

Solution. Given that $k = LCM\ \{m, n\}$ and $\gcd(m, n) = 1$. Therefore, $k = mn$.

Put $c = ab \in G$. Then $c^k = c^{mn} = (ab)^{mn} = (a^m)^n (b^n)^m$ [as G is abelian] $= e$.

Thus, $o(c) \leq k$. Let $o(c) = t$.

Then $t \leq mn$.

Now, $e = c^t = a^t b^t \Rightarrow a^t = b^{-t} \Rightarrow a^{mt} = b^{-mt}$

But $a^{mt} = e \Rightarrow b^{-mt} = e \Rightarrow b^{mt} = e$.

$o(b) = n$ implies $n|mt$. Therefore, $n|t$ as $\gcd(m, n) = 1$.

Similarly, it can be proved that $m|t$. Thus $mn \leq t$.

Hence, $mn = t$.

Thus it is shown that $o(c) = mn$.

Exercise

1. Let $G = H \times K$. Show that G is abelian if and only if both H, K are abelian.
2. Let $G = H \times K$, where G is a finite group. Show that $o(G) = o(H)o(K)$.
3. Show that the order of (8,4,10) in the group $\mathbb{Z}_{12} \times \mathbb{Z}_{60} \times \mathbb{Z}_{24}$ is 60.
4. If a group G is the internal direct product of its subgroups H, K, then $H \simeq G/K$ and $\frac{G}{H} \simeq K$.
   [**Hint :** For any $x \in G$, $x = hk$, $h \in H, k \in K$. Define $f : G \to H$, $g: G \to K$ by $f(x) = h$,
   $g(x) = k$. Show that f and g are homomorphisms onto H and K respectively with $kerf = K$ and $kerg = H$]
5. If $G = \langle a \rangle$ be any cyclic group of order , where $\gcd(m, n) = 1$. Let H and K be its subgroups of orders m and n respectively. Show that $G = H \times K$.
6. Let $G_1 = \langle a \rangle$, $G_2 = \langle b \rangle$ be two cyclic groups of order m and respectively such that $\gcd(m, n) > 1$. Show that $G = G_1 \times G_2$ is an abelian group of order mn, which is not cyclic.
7. Let G be a direct product of two subgroups, each of which is a cyclic group of order 5. Show that G cannot be cyclic.
8. Let G be a finite group having at least three elements in which $a^2 = e, \forall a \in G$. Show that G is internal direct product of a finite number of subgroups each of order 2 and $o(G) = 2^n$ for some $n \geq 2$.
9. If $Z(G)$ denotes the centre of a group G, then prove that $Z(G \times H) = Z(G) \times Z(H)$, G, H being groups. Hence deduce that $G \times H$ is abelian if and only if G and H are abelian.
10. Let N be a normal subgroup of a group G. If $G = H \times K$ where H and K are subgroups of G then prove that either N is abelian or N intersects H or K non-trivially.
11. If M and N are normal subgroups of a group G then show that $G/(M \cap N) \simeq G/M \times G/N$.
    [**Hint :** Define $f : G \to G/M \times G/N$ by $f(g) = (gM, gN)$.]
12. Let G be a group and $H = \{(g, g) : g \in G\}$. Show that H is a subgroup of $G \times G$. Further, H is a normal subgroup of $G \times G$ if and only if G is abelian.

Unit-2

Linear Algebra II

2.1 LINEAR ALGEBRA II

It is assumed that readers have sufficient knowledge of Vector spaces and its basis or dimension. But one thing should be noted that in the earlier part (see Linear Algebra I) no concepts of length, angle and distance were introduced. In geometry, we have idea of *dot product of vectors*. Keeping that in mind, let us introduce a new concept *inner product* on a vector space.

2.1.1 Definition. An **inner product** on a vector space V over a field F is a map $<,>: V \times V \to F$ satisfying the following properties : For $\alpha, \beta, \gamma \in V$ and $c \in F$,

(i) $<\alpha+\beta, \gamma> = <\alpha, \gamma> + <\beta, \gamma>$
(ii) $<c\alpha, \beta> = c<\alpha, \beta>$
(iii) $\overline{<\alpha, \beta>} = <\beta, \alpha>$ where the bar denotes complex conjugation
(iv) $<\alpha, \alpha> > 0$ if $\alpha \neq \theta$.

Conditions (i) and (ii) simply require that the inner product be linear in the first component. It is also to be observed that conditions (i), (ii), and (iii) together imply

$$\begin{aligned}<\alpha, c\beta+d\gamma> &= \overline{<c\beta+d\gamma, \alpha>}\\ &= \overline{c<\beta, \alpha> + d<\gamma, \alpha>}\\ &= \bar{c}\,\overline{<\beta, \alpha>} + \bar{d}\,\overline{<\gamma, \alpha>}\\ &= \bar{c}<\alpha, \beta> + <\alpha, \gamma>\end{aligned}$$

2.1.2 Definition. A complex vector space V together with a complex inner product defined on it, is called a **Unitary space.**

2.1.3 Example. On F^n there is an inner product which we call the **standard inner product.**

If $\alpha = (a_1, a_2, \dots, a_n)$ and $\beta = (b_1, b_2, \dots, b_n)$ in F^n, let us define

$$<\alpha, \beta> = \sum_{i=1}^{n} a_i \overline{b_i}$$

If $\gamma = (c_1, c_2, \ , c_n) \in F^n$, we have,

(i) $<\alpha+\beta, \gamma> = \sum_{i=1}^{n}(a_i+b_i)\overline{c_i} = \sum_{i=1}^{n} a_i\overline{c_i} + \sum_{i=1}^{n} b_i\overline{c_i}$

$$= <\alpha, \gamma> + <\beta, \gamma>$$

(ii) $<c\alpha, \beta> = \sum_{i=1}^{n} ca_i\overline{b_i} = c\sum_{i=1}^{n} a_i\overline{b_i} = c<\alpha, \beta>$

(iii) $\overline{<\alpha, \beta>} = \overline{\sum_{i=1}^{n} a_i\overline{b_i}} = \sum_{i=1}^{n} \overline{a_i\overline{b_i}} = \sum_{i=1}^{n} \overline{a_i}b_i = \sum_{i=1}^{n} b_i\overline{a_i} = <\beta, \alpha>$

(iv) If $\alpha \neq \theta$ at least one of a_i's is non zero. So,

$$<\alpha, \alpha> = \sum_{i=1}^{n} a_i\overline{a_i} = \sum_{i=1}^{n} |a_i|^2 > 0$$

Thus, properties of inner product are satisfied.

If $F = \mathbb{R}$, the conjugations are not needed, that is, in that case, we define

$$<\alpha, \beta> = a_1b_1 + a_2b_2 + \cdots + a_nb_n$$

This standard inner product is usually called the *dot product* and is denoted by $\alpha . \beta$ instead of $<\alpha, \beta>$.

We wish to define real inner product in a separate way.

2.1.4 Definition. A **real inner product** on a vector space V over $\mathbb{R}$ is a map $<,>: V \times V \to \mathbb{R}$ satisfying the following properties : For $\alpha, \beta, \gamma \in V$ and $c \in \mathbb{R}$

(i) $<\alpha+\beta, \gamma> = <\alpha, \gamma> + <\beta, \gamma>$ and
$<\alpha, \beta+\gamma> = <\alpha, \beta> + <\alpha, \gamma>$
(ii) $<\alpha, \beta> = <\beta, \alpha>$
(iii) $<c\alpha, \beta> = c<\alpha, \beta>$
(iv) $<\alpha, \alpha> \geq 0$ and $<\alpha, \alpha> = 0$ if and only if $\alpha = \theta$.

Now, $(V, <,>)$ is called an **inner product space.**

2.1.5 Definition. A real vector space V together with a real product defined on it, is called a **Euclidean space.**

Convention : Unless specified otherwise the inner product on $\mathbb{R}^n$ will be assumed to be the dot product.

2.1.6 Examples.

1. For $u, v \in \mathbb{R}^2$ where $u = (u_1, u_2)$, $v = (v_1, v_2)$, let us define

$$<u, v> = v_1(u_1 + 2u_2) + v_2(2u_1 + 5u_2)$$

It is easy to verify that this product satisfies the properties (i), (ii) and (iii).

Now, $< u,u > = u_1(u_1+2u_2)+u_2(2u_1+5u_2)$

$$= u_1^2+4u_1u_2+5u_2^2$$

$$= (u_1+2u_2)^2+u_2^2$$

Clearly, $< u,u > \geq 0$ and $< u,u > = 0$ if and only if $(u_1+2u_2)^2 = 0$ and $u_2^2 = 0$, that is, if and only if $u_1 = 0, u_2 = 0$. Thus, $< u,u > = 0$ if and only if $u = \theta$.

Thus, $\mathbb{R}^2$ becomes a Euclidean space under this inner product.

2. Let us consider the vector space $C[0,1]$, the space of all real valued continuous functions defined on $[0,1]$, over $\mathbb{R}$ and for $f,g \in C[0,1]$, define

$$< f,g > = \int_0^1 f(t)g(t)dt$$

It is easy to check that conditions (i), (ii) and (iii) are verified. To check the last one, we have

$$< f,f > = \int_0^1 [f(t)]^2\, dt \geq 0$$

Now, $< f,f > = 0$ if and only if $f = 0$ (follows from Analysis).

Thus, $(C[0,1], <,>)$ is an inner product space.

2.1.7 Definition. Let $(V, <,>)$ be an inner product space. The **length** or **norm** of a vector $v \in V$, denoted by $\|v\|$, is defined by $\|v\| = +\sqrt{< v,v >}$, that is, the norm of a vector v is the positive square root of the non-negative number $< v,v >$.

For example, if $v = (v_1, v_2, \ldots, v_n) \in F^n$ with standard inner product, then

$$\|v\| = +\sqrt{< v,v >} = +\sqrt{v_1\overline{v_1}+v_2\overline{v_2}+\cdots+v_n\overline{v_n}}$$
$$= \sqrt{|v_1|^2+|v_2|^2+\cdots+|v_n|^2}$$

If the reader feels uncomfortable with abstract inner product spaces, he may be advised to assume that the inner product space is $\mathbb{R}^n$ with the dot product introduced above.

2.1.8 Theorem. Let V be a real inner product space and $v \in V$. Then

(i) $\|v\| \geq 0$ and $\|v\| = 0$ if and only if $v = \theta$.

(ii) $\|cv\| = |c|\|v\|$ for $c \in \mathbb{R}$

Furthermore, for any non-null vector $v \in V$, there is a vector $u \in V$ such that $\|u\| = 1$ and $v = \|v\|u$. This u is called the unit vector along v.

Proof.

(i) Since, $\|v\| = +\sqrt{< v,v >}$, we have, $\|v\| \geq 0$ and

$\|v\| = 0$ if and only if $< v,v > = 0$ if and only if $v = \theta$.

(ii) We have,

$\|cv\|^2 = < cv,cv > = c < v,cv > = c < cv,v > = c^2 < v,v > = c^2\|v\|^2$

Hence, $\|cv\| = |c|\|v\|$.

For last part, take $u = \frac{v}{\|v\|}$. Then $\|u\| = \frac{\|v\|}{\|v\|} = 1$ and $v = \|v\|u$.

Hence proved.

One thing is to be noted: *working with norms squared is usually easier than working directly with norms.*

2.1.9 Cauchy-Schwarz's inequality

Let V be an inner product space over F. Then for $u, v \in V$,

$$|< u,v >| \leq \|u\|.\|v\|$$

Proof. If $v = \theta$, then $|< u,\theta >| = 0 = \|u\|.\|\theta\|$. So, the proof is done.

Let $v \neq \theta$. For any $c \in F$, we have,

$$0 \leq \|u-cv\|^2 = < u-cv, u-cv >$$

$$= < u,u-cv > \; - c < v,u-cv >$$

$$= < u,u > - \bar{c} < u,v > - c < v,u > + c\,\bar{c} < v,v >$$

Taking $c = \frac{<u,v>}{<v,v>}$, we have, $\bar{c} = \frac{\overline{<u,v>}}{<v,v>}$ and thus,

$$0 \leq \|u-cv\|^2 = \|u\|^2 - \frac{\overline{< u,v >}}{< v,v >} < u,v >$$

$$- \frac{< u,v >}{< v,v >} < v,u > \; + \frac{< u,v >\overline{< u,v >}}{< v,v >.< v,v >}.< v,v >$$

i.e.

$$0 \leq \|u\|^2 - \frac{(< u,v >)\overline{< u,v >}}{\|v\|^2} [as < v,u > = \overline{< u,v >}]$$

Therefore,

$$0 \leq \|u\|^2 - \frac{|< u,v >|^2}{\|v\|^2}$$

i.e. $$|< u,v >|^2 \leq \|u\|^2\|v\|^2$$

Hence,

$$|< u,v >| \leq \|u\|\|v\|.$$

Note. If we take V as Euclidean space then no conjugation required, that is, $c =$ and $< u,v > = < v,u >$ lead to the conclusion.

If we take $V = \mathbb{R}^n$ with standard inner product, then for $a, b \in \mathbb{R}^n$ where $a =$ $(a_1, a_2, .., a_n)$ and $b = (b_1, b_2, ..., b_n)$, using Cauchy-Schwarz's inequality, we get

$$|< a,b >|^2 \leq \|a\|^2\|b\|^2$$

That is,

$$(a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2 \leq (a_1^2 + a_2^2 + \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2)$$

which is a well-known result in Classical Algebra.

2.1.10 Triangle Inequality

Let V be an inner product space over a field F. Then for all $u, v \in V$

$$\|u + v\| \leq \|u\|\|v\|.$$

Proof. We have,

$$\|u + v\|^2 = < u + v, u + v >$$

$$= < u,u > + < u,v > + < v,u > + < v,v >$$

$$= \|u\|^2 + < u,v > + \overline{< u,v >} + \|v\|^2$$

$$= \|u\|^2 + 2\, Real < u,v > + \|v\|^2$$

$$\leq \|u\|^2 + 2|< u,v >| + \|v\|^2$$

$$\leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \quad [by\ Cauchy\ Schwarz's\ inequality]$$

$$= (\|u\| + \|v\|)^2$$

Hence, $\|u + v\| \leq \|u\| + \|v\|$.

We have learnt in vector algebra for $\mathbb{R}^2$ or $\mathbb{R}^3$ that $< u,v > = \|u\|\|v\| \cos\theta$ where $\theta (0 \leq \theta \leq \pi)$ denotes the angle between u and v. We also know that two non-zero vectors u and v are perpendicular if and only if $cos\theta = 0$, that is, if and only if $< u,v > = 0$. Then can we generalize the notion of perpendicularity to arbitrary inner product spaces? Let's try.

2.1.11 Definition. Let V be an inner product space. Vectors u and v in V are **orthogonal (perpendicular)** if $< u,v > = 0$.

A subset $S = \{v_1, v_2, ..., v_n\}$ of V is **orthogonal** if $< v_i, v_j > = 0$ for $i \neq j$.

A vector v in V is a unit vector if $\|v\| = 1$. If v is any non-null vector in V, then $\frac{v}{\|v\|}$ is a unit vector.

A subset $S = \{v_1, v_2, ..., v_n\}$ is said to be **orthonormal** if $< v_i, v_j > = 0$ for $i \neq j$ and $< v_i, v_j > = 1$ for $i = j$.

Look, the null vector θ is orthogonal to any vector. Furthermore, θ is the only vector in V which is orthogonal to itself. So, an orthogonal set may contain θ but an orthonormal set cannot. An orthonormal set is an orthogonal set consisting of unit vectors only.

2.1.12 Example. In F^3, the set $S = \{(1,1,0), (1,-1,0), (-1,1,2)\}$ is orthogonal as

$$< (1,1,0), (1,-1,0) > = 1.1 + 1(-1) + 0.0 = 0$$

$$< (1,1,0), (-1,1,2) > = 1(-1) + 1.1 + 0.2 = 0$$

and $< (1,-1,0), (-1,1,2) > = 1(-1) + (-1)1 + 0.2 = 0$

But S is not orthonormal as $\|(1,-1,0)\| = \sqrt{1^2 + (-1)^2 + 0^2} = \sqrt{2}$. However, we can obtain an orthonormal set from S by normalizing the vectors, as

$$\left\{\frac{1}{\sqrt{2}}(1,1,0), \frac{1}{\sqrt{2}}(1,-1,0), \frac{1}{\sqrt{6}}(-1,1,2)\right\}$$

2.1.13 Pythagoras Theorem

If u, v be two orthogonal vectors in a Euclidean space V, then

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$

Proof. We have,

$$\|u + v\|^2 = < u + v, u + v >$$

$$= < u,u > + < u,v > + < v,u > + < v.v >$$

$$= \|u\|^2 + \|v\|^2 [as < u,v > = < v,u > = 0]$$

Hence proved.

Note. If we take $V = \mathbb{R}^2$, then theorem 2.1.13 matches with the 2500 years old Pythagoras theorem what you read in school level geometry.

2.1.14 Parallelogram law

If u, v be any two vectors in a Euclidean space V, then

$$\|u+v\|^2 + \|u-v\|^2 = 2\,\|u\|^2 + 2\,\|v\|^2$$

Proof. We have,

$$\|u+v\|^2 = <u+v, u+v> = <u,u> + <u,v> + <v,u> + <v,v>$$
$$= \|u\|^2 + 2 <u,v> + \|v\|^2 \ldots (1)$$

Similarly,

$$\|u-v\|^2 = \|u\|^2 - 2 <u,v> + \|v\|^2 \ldots (2)$$

By (1)and (2), we have,

$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2 \blacksquare$$

2.1.15 Theorem. An orthogonal set of non-null vectors is linearly independent.

Proof. Let S be an orthogonal set of non-null vectors in a given inner product space V over a field F.Let $v_1, v_2, .., v_k$ are distinct vectors in S such that for $c_1, c_2, .., c_n \in F$,

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = \theta$$

Then for $k \in \{1,2,..,n\}$, we have,

$$< c_1 v_1 + c_2 v_2 + \cdots + c_n v_n, v_k > = < \theta, v_k > = 0$$

$$=> c_1 < v_1, v_k > + c_2 < v_2, v_k > + \cdots + c_k < v_k, v_k > + \cdots + c_n < v_n, v_k > = 0$$

$$=> c_k < v_k, v_k > = 0 \, [\, as < v_i, v_k > = 0 \;\; if\; i \neq k]$$

Since, $< v_k, v_k > \neq 0$, we have, $c_k = 0$ for $k = 1,2,..,n$.

Hence, S is linearly independent.

Corollary. Any orthonormal set of vectors in an inner product space is linearly independent.

Let $u, v \in V$ where V is an inner product space over a field F. Let $v \neq \theta$. Let us try to write u as a sum of some scalar multiple of v and a vector w which is orthogonal to v. Let $a \in F$. We wish to express $u = av + w$

where$w$ is orthogonal to v. In other words, $<w, v> = 0$, that is,

$$< u - av, v > = 0, \qquad i.e. \; < u, v > - a < v, v > = 0$$

Thus,

$$a = \frac{< u, v >}{\|v\|^2}$$

By this choice of$a$, we have,

$$u = \frac{< u, v >}{\|v\|^2} v + \left(u - \frac{< u, v >}{\|v\|^2} v \right)$$

This technique, known as **orthogonal decomposition,** can be used to prove Cauchy- Schwarz inequality for any inner product space V. Theorem 2.1.9 proved this result for Euclidean spaces only. Now, we are going to prove it for general case.

2.1.16 Theorem. If $u, v \in V$, then

$$|< u, v >| \leq \|u\|\|v\|$$

This inequality is an equality if and only if one of u, v is a scalar multiple of other.

Proof. Let $u, v \in V$. If $v = \theta$, then $|< u, \theta >| = 0 = \|u\|^2\|\theta\|^2$.

So, let $v \neq \theta$. Let us consider the orthogonal decomposition

$$u = \frac{< u, v >}{\|v\|^2} v + w$$

where$w \in V$ and $< w, v > = 0$.

By the Pythagorean theorem,

$$\|u\|^2 = \left\| \frac{< u, v >}{\|v\|^2} v \right\|^2 + \|w\|^2$$

$$=> \|u\|^2 \geq \frac{|< u, v >|^2}{\|v\|^2} [\, as\; \|w\|^2 \geq 0]$$

$$=> |< u, v >|^2 \leq \|u\|^2 \|v\|^2$$

Taking the square root,

$$|< u, v >| \leq \|u\|\|v\|.$$

Equality holds if and only if $\|u\|^2 = \left\| \frac{<u,v>}{\|v\|^2} v \right\|^2$, if and only if, $\|w\|^2 = 0$, if and only if $w = \theta$.

Thus, equality holds if and only if $u = \frac{<u,v>}{\|v\|^2} v$, that is, if and only if u is a scalar multiple of $v \blacksquare$

By theorem 2.1.15 and its corollary it is clear that every orthonormal set of vectors is linearly independent. Let V be an inner product space with $\dim V = n$. Then any orthonormal set of vectors in V containing n elements is a basis of V.

For example, standard basis of F^n. In general, if $\{v_1, v_2, .., v_n\}$ is a basis of V, then any element v of V can be expressed as

$$v = a_1v_1 + a_2v_2 + \cdots + a_nv_n$$

for some choice of scalars $a_1, a_2, \dots, a_n$. But it is not so easy to find these scalars. However, this is easy for orthonormal basis. The next theorem will support it.

2.1.17 Theorem. Let $\{e_1, e_2, \dots, e_n\}$ be an orthonormal basis of $V$and $v \in V$. Then

$$v = <v, e_1> e_1 + <v, e_2> e_2 + \cdots + <v, e_n> e_n$$

and $\quad \|v\|^2 = |<v, e_1>|^2 + |<v, e_2>|^2 + \cdots + |<v, e_n>|^2$

Proof. Since, $\{e_1, e_2, .., e_n\}$ is a basis of V, we have,

$$v = a_1e_1 + a_2e_2 + \cdots + a_ne_n$$

where$a_1, a_2, .., a_n$ are scalars.

Now, $<v, e_j> = a_1 <e_1, e_j> + a_2 <e_2, e_j> + \cdots + a_j <e_j, e_j>$
$+ \cdots + a_n <e_n, e_j>$

$$= a_j (\, as <e_i, e_j> = 0 \, for \, i \neq j \, and <e_j, e_j> = 1)$$

for $j = 1,2, \dots, n$.

Thus, $\quad v = <v, e_1> e_1 + <v, e_2> e_2 + \cdots + <v, e_n> e_n$

By repeated application of Pythagorean theorem, we have

$$\|v\|^2 = |<v, e_1>|^2 + |<v, e_2>|^2 + \cdots + |<v, e_n>|^2 \blacksquare$$

I think, now the usefulness of an orthonormal basis is clear. So, can we convert a given basis of V to an orthonormal basis of V? The answer is given by the following theorem that includes a process, known as **Gram- Schmidt orthonormalisation process.**

2.1.18 Theorem . Let V be any finite dimensional inner product spacewith $dimV = n \geq 1$. Then V has an orthonormal basis.

Proof. Let $\{v_1, v_2, .., v_n\}$ be a basis of V. We first try to produce an orthogonal basis of V.

Let $u_1 = v_1$.

We define, $u_2 = v_2 - \frac{<v_2, u_1>}{\|u_1\|^2} u_1$

Then,$<u_2, u_1> = <v_2 - \frac{<v_2,u_1>}{\|u_1\|^2} u_1, u_1>$

$$= <v_2, u_1> - \frac{<v_2, u_1>}{\|u_1\|^2} <u_1, u_1> = 0$$

Moreover, $\quad u_2 \neq \theta$. Because, if $u_2 = \theta$ then $v_2 = \frac{<v_2,u_1>}{\|u_1\|^2} v_1$ shows that $\{v_1, v_2\}$ is linearly dependent and hence $\{v_1, v_2, \dots, v_n\}$ is linearly dependent which is not the case. Let

$$u_3 = v_3 - \frac{<v_3, u_1>}{\|u_1\|^2} u_1 - \frac{<v_3, u_2>}{\|u_2\|^2} u_2$$

Now,

$$<u_3, u_1> = <v_3, u_1> - \frac{<v_3, u_1>}{\|u_1\|^2} <u_1, u_1> - \frac{<v_3, u_2>}{\|u_2\|^2} <u_1, u_2> = 0$$

Similarly, $\quad <u_3, u_2> = 0$

$u_3 \neq \theta$. For otherwise, v_3 is a linear combination of u_1 and u_2 and hence a linear combination of $v_1$and v_2. This implies $\{v_1, v_2, v_3\}$ is linearly dependent, a contradiction.

Proceeding as above by induction, define,

$$u_k = v_k - \frac{<v_k, u_1>}{\|u_1\|^2} u_1 - \frac{<v_k, u_2>}{\|u_2\|^2} u_2 - \cdots - \frac{<v_k, u_{k-1}>}{\|u_{k-1}\|^2} u_{k-1}$$

Then, $<u_k, u_i> = 0$ for $1 \leq i \leq k-1$ and as before $u_k \neq \theta$.

This process terminates after n steps and thus we have produced an orthogonal basis $\{u_1, u_2, \dots, u_n\}$of V.

Hence, we get an orthonormal basis $\{e_1, e_2, \dots, e_n\}$ of V where $e_i = \frac{u_i}{\|u_i\|}$ for $\quad i = 1,2, \dots, n$.

Therefore, proof is complete.

2.1.19Example. We wish to form an orthonormal basis of $\mathbb{R}^3$ from a given basis $\{(-1,0,1), (1,-1,0), (0,0,1)\}$.

Let $u_1 = (-1,0,1)$. Then

$$u_2 = u_2 - \frac{<u_2, u_1>}{\|u_1\|^2} u_1$$

$$= (1, -1,0) - \frac{<(1, -1,0), (-1,0,1)>}{\|(-1,0,1)\|^2} (-1,0,1)$$

$$= (1,-1,0) - \frac{1(-1)+(-1)0+0.1}{2}(-1,0,1)$$

$$= (1,-1,0) + \frac{1}{2}(-1,0,1) = (1,-1,0) + \left(-\frac{1}{2},0,\frac{1}{2}\right)$$

$$= \left(\frac{1}{2},-1,\frac{1}{2}\right)$$

So, $u_2 = \left(\frac{1}{2},-1,\frac{1}{2}\right)$ and $\|u_2\|^2 = \frac{1}{4}+1+\frac{1}{4} = \frac{3}{2}$

Now, $u_3 = u_3 - \frac{<u_3,u_1>}{\|u_1\|^2}u_1 - \frac{<u_3,u_2>}{\|u_2\|^2}u_2$

$$= (0,0,1) - \frac{<(0,0,1),\ (-1,0,1)>}{2}(-1,0,1) - \frac{<(0,0,1),\left(\frac{1}{2},-1,\frac{1}{2}\right)>}{\frac{3}{2}}\left(\frac{1}{2},-1,\frac{1}{2}\right)$$

$$= (0,0,1) - \frac{1}{2}(-1,0,1) - \frac{2}{3}\cdot\frac{1}{2}\left(\frac{1}{2},-1,\frac{1}{2}\right)$$

$$= (0,0,1) - \left(-\frac{1}{2},0,\frac{1}{2}\right) - \left(\frac{1}{6},-\frac{1}{3},\frac{1}{6}\right)$$

$$= \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$$

and $\|u_3\|^2 = \frac{1}{9}+\frac{1}{9}+\frac{1}{9} = \frac{1}{3}$

Therefore, $\left\{(-1,0,1),\left(\frac{1}{2},-1,\frac{1}{2}\right),\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)\right\}$ is an orthogonal basis of $\mathbb{R}^3$ and

$$\left\{\frac{1}{\sqrt{2}}(-1,0,1),\ \sqrt{\frac{2}{3}}\left(\frac{1}{2},-1,\frac{1}{2}\right),\sqrt{3}\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)\right\}$$

is an orthonormal basis of $\mathbb{R}^3$.

Solved Examples :

1. **Examine whether each of the following is an inner product on the given vector spaces**
   (i) $<(a,b),(c,d)> = ac - bd$ on $\mathbb{R}^2$
   (ii) $<A,B> = tr(A+B)$ on $M_{2\times 2}(\mathbb{R})$
   (iii) $<u,v> = |u_1v_1+u_2v_2+u_3v_3|$ **where** $u=(u_1,u_2,u_3), v=(v_1,v_2,v_3)$ **belongs to the vector space** $\mathbb{R}^3$.

Solution.

(i) This is not an inner product on $\mathbb{R}^2$ as $<(1,1),(1,1)> = 1.1 - 1.1 = 0$ but $(1,1)$ is not the null vector in $\mathbb{R}^2$.

(ii) This is not an inner product as for $A = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$

$<A,A> = tr\,(A+A) = tr\begin{pmatrix}2 & 0\\ 0 & -2\end{pmatrix} = 2-2 = 0$ but A is not the null matrix in $M_{2\times 2}(\mathbb{R})$.

(iii) Let $u=(u_1,u_2,u_3), v=(v_1,v_2,v_3) \in \mathbb{R}^3$ and $-1 \in \mathbb{R}$. Then

$$<(-1)u,v> = |-u_1v_1-u_2v_2-u_3v_3| = |u_1v_1+u_2v_2+u_3v_3|$$

But, $(-1)<u,v> = -|u_1v_1+u_2v_2+u_3v_3|$.

Therefore, $<(-1)u,v> \neq (-1)<u,v>$.

Hence, $<,>$ is not an inner product.

2. **Let $V=C[0,1]$ be the vector space consisting of all real valued continuous functions defined on $[0,1]$ and define $<,>: V\times V \to \mathbb{R}$ by $<f,g> = \int_0^1 f(t)g(t)dt$. Examine whether $<f,g>$ is an inner product.**

Solution. Let $f,g,h \in V$ and $c \in \mathbb{R}$. Then

(i) $<f,f> = \int_0^1 [f(t)]^2 dt \geq 0$ as integral of a non-negative function defined on $[0,1]$ is always non-negative and $<f,f> = 0$ if and only if $\int_0^1 [f(t)]^2 dt = 0$ if and only if $f(t) = 0$ for all $t \in [0,1]$.
Thus, $<f,f> = 0$ if and only if f is a zero function.

(ii) $<f,g> = \int_0^1 f(t)g(t)dt = \int_0^1 g(t)f(t)dt = <g,f>$.

(iii) $<cf,g> = \int_0^1 cf(t)g(t)dt = c\int_0^1 f(t)g(t)dt = c<f,g>$.

(iv) $<f+g,h> = \int_0^1 [(f+g)(t)h(t)dt = \int_0^1 [f(t)+g(t)]h(t)dt$

$$= \int_0^1 f(t)h(t)dt + \int_0^1 g(t)h(t)dt$$

$$= <f,h> + <g,h>$$

Hence, $(V,<,>)$ is an inner product space.

3. Let $x = (2, 1+i, i)$ and $y = (2-i, 2, 1+2i)$ be vectors in $\mathbb{C}^3$. Compute $< x, y >, \|x\|, \|y\|$ and $\|x+y\|$. Then verify both the Cauchy Schwarz's inequality and the triangle inequality.

Solution. We know, $< x, y > = x_1\overline{y_1} + x_2\overline{y_2} + x_3\overline{y_3}$

if $x = (x_1, x_2, x_3), y = (y_1, y_2, y_3)$.

Thus, $< x, y > = 2(2+i) + (1+i)2 + i(1-2i)$

$$= 4 + 2i + 2 + 2i + i + 2 = 8 + 5i.$$

$$\|x\|^2 = < x, x > = 2.2 + (1+i)(1-i) + i(-i) = 4 + 1 + 1 + 1 = 7$$

So, $\|x\| = \sqrt{7}$.

Again,

$$\|y\| = \sqrt{(2-i)(2+i) + 2.2 + (1+2i)(1-2i)} = \sqrt{4+1+4+1+4} = \sqrt{14}$$

$$\|x+y\| = \|(4-i, 3=i, 1+3i)\|$$
$$= \sqrt{(4-i)(4+i) + (3-i)(3+i) + (1+3i)(1-3i)} = \sqrt{37}$$

Now, $|< x, y >| = \sqrt{8^2 + 5^2} = \sqrt{89} \le \sqrt{7}\sqrt{14} = \sqrt{98} = \|x\|\|y\|$.

Hence, Cauchy Schwarz's inequality is verified.

Clearly, $\sqrt{7} + \sqrt{14} \ge \sqrt{37}$ i.e. $\|x\| + \|y\| \ge \|x+y\|$.

4. Suppose $u, v \in V$ where V is an inner product space over a field F. Prove that $< u, v > = 0$ if and only if $\|u\| \le \|u + av\|$ for all $a \in F$.

Solution. Let us suppose first $< u, v > = 0$. Then by Pythagorean theorem, we get,

$$\|u + av\|^2 = \|u\|^2 + \|av\|^2, \qquad \forall a \in F.$$

Thus,

$$\|u + av\|^2 \ge \|u\|^2 \;(as\; \|av\|^2 \ge 0), \qquad i.e.\quad \|u\| \le \|u + av\|.$$

Conversely, let $\|u\| \le \|u + av\|$ for all $a \in F$. Squaring the inequality, we get,

$$\|u\|^2 \le \|u + av\|^2 = < u + av, u + av >$$
$$= < u, u > + < u, av > + < av, u > + < av, av >$$
$$= \|u\|^2 + \bar{a} < u, v > + a\,\overline{< u, v >} + a\bar{a} < v, v >$$
$$= \|u\|^2 + 2Re\,\bar{a} < u, v > + |a|^2\|v\|$$

Thus, for all $a \in F$, we get

$$-2Re\,\bar{a} < u, v > \le |a|^2\|v\|^2.$$

In particular, taking $a = -t < u, v >$ where $t > 0$, we have,

$$2t\,|< u, v >|^2 \le t^2 |< u, v >|^2\|v\|^2$$

That is,

$$2|< u, v >|^2 \le t\,|< u, v >|^2\|v\|^2 [\,as\; t > 0].$$

for all $t > 0$.

If $v = \theta$ then $< u, v > = 0$. If $v \ne \theta$, taking $t = \frac{1}{\|v\|^2}$, we get,

$$2\,|< u, v >|^2 \le |< u, v >|^2$$

which is possible only when $< u, v > = 0$.

Hence the problem.

5. Suppose $u, v \in V$ are such that $\|u\| = 3$, $\|u + v\| = 4$, $\|u - v\| = 6$. Find the value of $\|v\|$.

Solution. By parallelogram law, we have,

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2.$$

Putting the given values, we get $\|v\| = 17$.

6. Let β be a basis for a finite dimensional inner product space. Then prove that

(i) If $< x, z > = 0$ for all $z \in \beta$ then $x = \theta$.

(ii) If $< x, z > = < y, z >$ for all $z \in \beta$ then $x = y$.

Solution. (i) Let $x \in V$. Let $\beta = \{z_1, z_2, \dots, z_n\}$. Thus, there exist scalars $c_1, c_2, \dots, c_n$ such that

$$x = c_1z_1 + c_2z_2 + \dots + c_nz_n$$

Now,

$$< x, x > = < x,\; c_1z_1 + c_2z_2 + \dots + c_nz_n >$$
$$= \overline{c_1} < x, z_1 > + \overline{c_2} < x, z_2 > + \dots + \overline{c_n} < x, z_n >$$
$$= 0\; [\,as < x, z_i > = 0, \qquad i = 1,2, \dots, n]$$

Therefore, $< x, x > = 0$ which implies $x = \theta$.

(ii) $<x,z> = <y,z> \Rightarrow <x-y,z> = 0$ for all $z \in \beta$.

Hence, by (i), we have, $x - y = \theta$, i.e. $x = y$.

7. **Let $\{v_1, v_2, \ldots, v_n\}$ be an orthogonal set in V, and let $a_1, a_2, \ldots, a_n$ be scalars. Then prove that n**

$$\left\|\sum_{i=1}^{n} a_i v_i\right\|^2 = \sum_{i=1}^{n} |a_i|^2 \|v_i\|^2.$$

Solution. We have,

$$\left\|\sum_{i=1}^{n} a_i v_i\right\|^2 = <\sum_{i=1}^{n} a_i v_i , \sum_{i=1}^{n} a_i v_i>$$

$$= \sum_{i=1}^{n} a_i \bar{a}_i < v_i, v_i > [as < v_i, v_j >= 0 \, for \, i \neq j]$$

$$= \sum_{i=1}^{n} |a_i|^2 \|v_i\|^2$$

8. **Prove that if V is an inner product space, then $|<x,y>| = \|x\|.\|y\|$ if and only if one of the vectors x or y is a scalar multiple of other, that is, $\{x,y\}$ is linearly dependent.**

Solution. If one of x or y is zero then both sides of the equality becomes zero and we can write $y = 0x$ or $x = 0y$.

Let $x \neq 0$. If $y = cx$ for some scalar c, then

$$|<x,y>| = |<x,cx>| = |\bar{c}<x,x>| = |c|\|x\|^2$$

and $\|x\|\|y\| = \|x\|\|cx\| = |c|\|x\|^2$

Thus, if $y = cx$ we have, $|<x,y>| = \|x\|\|y\|$.

Conversely, let $|<x,y>| = \|x\|\|y\|$.

Now, $\|cx - y\|^2 = <cx-y, cx-y>$

$$= <cx,cx> - <y,cx> - <cx,y> + <y,y>$$

$$= |c|^2\|x\|^2 - \bar{c}\,\overline{<x,y>} - c<x,y> + \|y\|^2$$

If we take $c = \frac{\overline{<x,y>}}{\|x\|^2}$, we get

$$\|cx-y\|^2 = \frac{|<x,y>|^2}{\|x\|^2} - \frac{|<x,y>|^2}{\|x\|^2} - \frac{|<x,y>|^2}{\|x\|^2} + \frac{|<x,y>|^2}{\|x\|^2}\left[as \ \|y\|^2 = \frac{|<x,y>|^2}{\|x\|^2}\right]$$

which shows that $\|cx - y\| = 0$. In other words, $cx - y = \theta$, that is, $y = cx$.

Hence proved.

9. **Let V be a unitary space with $dim\ V = n$ and let $u \in V$ be such that u is orthogonal to n linearly independent vectors of V. Then prove that$u = \theta$.**

Solution. Let $S = \{v_1, v_2, \ldots, v_n\}$ be a set of n linearly independent vectors of V, each member of which is orthogonal to u. Since, $\dim V = n$, it is clear that S is a basis for V. Thus, there exist scalars $c_1, c_2, \ldots, c_n$ such that

$$u = c_1v_1 + c_2v_2 + \cdots + c_nv_n$$

Now,
$$<u,u> = <c_1v_1 + c_2v_2 + \cdots + c_nv_n, u>$$
$$= c_1 <v_1,u> + c_2 <v_2,u> + \cdots + c_n <v_n,u>$$
$$= 0 \ [as < v_i, u >= 0 \ for \ i = 1,2,\ldots,n]$$

Thus, $<u,u> = 0$ which proves that $u = \theta$.

10. **Consider $\mathbb{R}^4$ with the standard inner product. Let W be the subspace of $\mathbb{R}^4$ consisting of all vectors which are orthogonal to both $u = (1,0,-1,1)$and $v = (2,3,-1,2)$. Find a basis for W.**

Solution. Let $w \in W$where $w = (x,y,z,t)$. Now,

$$<w,u> = 0 \Rightarrow <(x,y,z,t),(1,0,-1,1)> = 0 \Rightarrow x - z + t = 0 \ldots (1)$$

and
$$<w,v> = 0 \Rightarrow <(x,y,z,t),(2,3,-1,2)> = 0$$
$$\Rightarrow 2x + 3y - z + 2 = 0 \ldots (2)$$

By (1), $x = z - t$. Putting this value of x in (2)

$$2z - 2t + 3y - z + 2t = 0 \Rightarrow z = -3y \ldots (3)$$

Again, $x = z - t \Rightarrow x = -3y - t$. Hence,

$$w = (x,y,z,t) = (-3y-t, y, -3y, t) = y(-3,1,-3,0) + t(-1,0,0,1), \quad y,t \in \mathbb{R}.$$

Hence, $W = span\{(-3,1,-3,0),(-1,0,0,1)\}$.

Again, $(-3,1,-3,0)$ and $(-1,0,0,1)$ are linearly independent as no one is the scalar multiple of other. Therefore, $\{(-3,1,-3,0),(-1,0,0,1)\}$ is a basis for $W$and $\dim W = 2$.

11. Suppose $\{e_1, e_2, \dots, e_m\}$ is an orthonormal list of vectors in V. Let $v \in V$. Prove that $\|v\|^2 = |< v, e_1 >|^2 + |< v, e_2 >|^2 + \cdots + |< v, e_m >|^2$ if and only if $v \in span\{e_1, e_2, \dots, e_m\}$.

Solution. By the problem the set $\{e_1, e_2, \dots, e_m\}$ is linearly independent as it is orthonormal. Thus, it can be extended to an orthonormal basis, say $\{e_1, e_2, \dots, e_m, e_{m+1}, \dots, e_n\}$,for V. Hence, by theorem 2.1.17, a vector $v \in V$ can be written as

$$v = < v, e_1 > e_1 + \cdots + < v, e_m > e_m + < v, e_{m+1} > e_{m+1} + \cdots + < v, e_n > e_n$$

and

$$\|v\|^2 = |< v, e_1 >|^2 + \cdots + |< v, e_m >|^2 + |< v, e_{m+1} >|^2 + \cdots + |< v, e_n >|^2.$$

Thus, $\|v\|^2 = |< v, e_1 >|^2 + \cdots + |< v, e_m >|^2$

ifand only if $< v, e_{m+1} > = \cdots = < v, e_n > = 0$

i.e. if and only if $v = < v, e_1 > e_1 + \cdots + < v, e_m > e_m$

if and only if $v \in span\,\{v_1, v_2, \dots, v_m\}$.

12. Extend the set of vectors $\{(2, -3, 1), (1, 2, 4)\}$ to an orthogonal basis for the Euclidean space $\mathbb{R}^3$.

Solution. Let $v_1 = (2, -3, 1)$and $v_2 = (1, 2, 4)$. Let $v_3 = (0, 0, 1)$. Since,

$$\begin{vmatrix} 2 & -3 & 1 \\ 1 & 2 & 4 \\ 0 & 0 & 1 \end{vmatrix} = 7 \neq 0$$

we can say that $\{v_1, v_2, v_3\}$ is linearly independent in $\mathbb{R}^3$ and hence is a basis for $\mathbb{R}^3$ as $\dim \mathbb{R}^3 = 3$.

Now, $< v_1, v_2 > = 2.1 + (-3).2 + 1.4 = 0$. So, we take another vector w as

$$w = v_3 - c_1 v_1 - c_2 v_2$$

where$c_1 = \frac{< v_3, v_1 >}{\|v_1\|^2} = \frac{1}{14}$ and $c_2 = \frac{< v_3, v_2 >}{\|v_2\|^2} = \frac{4}{21}$. Thus,

$$w = (0,0,1) - \frac{1}{14}(2, -3, 1) - \frac{4}{21}(1,2,4) = \left(-\frac{1}{3}, -\frac{1}{6}, -\frac{3}{14}\right)$$

Hence, $\{v_1, v_2, w\}$ is an orthogonal basis for $\mathbb{R}^3$ which is an extension of the given set.

13. Use Gram-Schmidt process to convert the given basis $\{(1, 2, -2), (2, 0, 1), (1, 1, 0)\}$ of the Euclidean space $\mathbb{R}^3$ with the standard inner product into an orthogonal basis.

Solution. Let $v_1 = (1, 2, -2)$, $v_2 = (2, 0, 1)$, $v_3 = (1, 1, 0)$. Clearly,

$$< v_1, v_2 > = 1.2 + 2.0 + (-2).1 = 0$$

By Gram-Schmidt process, we take,

$$w = v_3 - \frac{< v_3, v_1 >}{\|v_1\|^2} v_1 - \frac{< v_3, v_2 >}{\|v_2\|^2} v_2$$

$$= (1,1,0) - \frac{1 + 2 + 0}{9}(1, 2, -2) - \frac{2}{5}(1,1,0)$$

$$= (1,1,0) - \frac{1}{3}(1, 2, -2) - \frac{2}{5}(1,1,0)$$

$$= \left(-\frac{2}{15}, \frac{1}{3}, \frac{4}{15}\right) = \frac{1}{15}(-2, 1, 4)$$

Thus, a converted orthogonal basis of $\mathbb{R}^3$ is given by $\left\{(1, 2, -2), (2, 0, 1), \frac{1}{15}(-2, 1, 4)\right\}$.

14. Apply Gram-Schmidt process to find an orthonormal basis for the Euclidean space $\mathbb{R}^3$with standard inner product that contains the vector $\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right)$.

Solution. We may take standard basis of $\mathbb{R}^3$as $\{(1,0,0), (0,1,0), (0,0,1)\}$. Since

$$\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right) = \frac{1}{\sqrt{2}}(1,0,0) - \frac{1}{\sqrt{2}}(0,1,0) + 0(0,0,1)$$

by replacement theorem, we can say that $\left\{\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right), (0,1,0), (0,0,1)\right\}$ is a basis for $\mathbb{R}^3$.

Now, we apply Gram-Schmidt process.

Let $v_1 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right)$. We take

$$v_2 = (0,1,0) - \frac{< (0,1,0), \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right) >}{1}\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right)$$

$$= (0,1,0) + \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right) = \left(\frac{1}{2}, \frac{1}{2}, 0\right)$$

and

$$v_3 = (0,0,1) - \frac{<(0,0,1),\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},0\right)}{1}\left(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}},0\right)$$
$$-\frac{<(0,0,1),\left(\frac{1}{2},\frac{1}{2},0\right)}{\frac{1}{2}}\left(\frac{1}{2},\frac{1}{2},0\right) = (0,0,1)$$

Hence, the required orthogonal basis for $\mathbb{R}^3$ is given by

$$\left\{\left(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}},0\right),\left(\frac{1}{2},\frac{1}{2},0\right),(0,0,1)\right\}$$

and corresponding orthonormal basis for $\mathbb{R}^3$ is given by

$$\left\{\left(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}},0\right),\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},0\right),(0,0,1)\right\} \blacksquare$$

15. Find an orthonormal basis for the row space of the matrix

$$A = \begin{pmatrix} 1 & 1 & 11 \\ 1 & 2 & 10 \\ 2 & 1 & 23 \end{pmatrix}.$$

Solution. Let us try to reduce the given matrix A to a row echelon matrix by applying elementary row operations on it as follows.

$$A \to \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & -1 & 0 & 1 \end{pmatrix} [R_2' = R_2 - R_1, R_3' = R_3 - 2R_1].$$

$$\to \begin{pmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} [R_1' = R_1 - R_2, R_3' = R_3 + R_2]$$

whichis in echelon form whose non-zero row vectors are $(1,0,1,2)$ and $(0,1,0,-1)$. These non-zero vectors form a basis for the row space of A.

Let $v_1 = (1,0,1,2)$. Using Gram-Schmidt process, we may take v_2 as

$$v_2 = (0,1,0,-1) - \frac{<(1,0,1,2),(0,1,0,-1)>}{\|(1,0,1,2)\|^2}(1,0,1,2)$$

$$= (0,1,0,-1) - \frac{(-2)}{6}(1,0,1,2) = \frac{1}{3}(1,3,1,-1)$$

Therefore, $\left\{(1,0,1,2),\frac{1}{3}(1,3,1,-1)\right\}$ is an orthogonal basis for the row space of A and the corresponding orthonormal basis for the row space of A is given by

$$\left\{\frac{1}{6}(1,0,1,2),\frac{1}{6\sqrt{3}}(1,3,1,-1)\right\} \blacksquare$$

16. In $\mathbb{R}^4$, let $U = span\{(1,1,0,0),(1,1,1,2)\}$. Find $u \in U$ such that $\|u-(1,2,3,4)\|$ is as small as possible.

Solution. Using Gram-Schmidt process, we get the orthonormal basis of U as $\{e_1,e_2\}$ where

$$e_1 = \left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},0,0\right), \quad e_2 = \left(0,0,\frac{1}{\sqrt{5}},\frac{2}{\sqrt{5}}\right)$$

Thus, the nearest point u to (1,2,3,4)is given by

$$u = <(1,2,3,4),e_1> e_1 + <(1,2,3,4),e_2> e_2$$
$$= \left(\frac{3}{2},\frac{3}{2},\frac{11}{5},\frac{22}{5}\right) [\,on\ simplification]$$

Hence the required point u for which $\|u-(1.2.3.4)\|$ is minimum, given by

$$u = \left(\frac{3}{2},\frac{3}{2},\frac{11}{5},\frac{22}{5}\right).$$

Exercise

1. Prove that if V is a real inner product space then
$$<u,v> = \frac{\|u+v\|^2 - \|u-v\|^2}{4}$$
for all $u,v \in V$.
2. Prove that if V is a complex inner product space then
$$<u,v> = \frac{\|u+v\|^2 - \|u-v\|^2 + \|u+iv\|^2 i - \|u-iv\|^2 i}{4}.$$
3. Prove that in an Euclidean vector space, two vectors u and v are orthogonal if and only if $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.
4. Show that the set of vectors $\{(2,-2,1),(1,2,-2),(2,1,2)\}$is an orthogonal basis for the inner product space $\mathbb{R}^3$ with standard inner product. Express $(1,2,3)$ as a linear combination of the aforesaid basis.
5. Use Gram-Schmidt process to convert the given basis $\{(1,1,0),(0,1,1),(1,0,1)\}$ of the Euclidean space $\mathbb{R}^3$ with the standard inner product into an orthonormal basis.
6. Apply Gram-Schmidt process to find an orthonormal basis for the Euclidean space $\mathbb{R}^3$with standard inner product that contains the vectors $\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},-\frac{1}{\sqrt{3}}\right)$ and $\left(\frac{2}{\sqrt{6}},\frac{-1}{\sqrt{6}},\frac{1}{\sqrt{6}}\right)$.
7. Find an orthonormal basis for the row space of the matrix
$$A = \begin{pmatrix} 1 & 1 & 10 \\ 2 & 3 & 11 \\ 1 & 2 & 31 \end{pmatrix}.$$

2.2 ORTHOGONAL COMPLEMENTS

Let V be an inner product space and U be a subset of V. We wish to find subset of V, denoted by, $U^{\perp}$, such that each vector of $U^{\perp}$ is orthogonal to vectors of U. Formal definition given below.

2.2.1 Definition. Let V be a vector space and $U \subseteq V$. Define

$$U^{\perp} = \{x \in V : < x, u >= 0 \,, \forall u \in U\}.$$

$U^{\perp}$is called the **orthogonal complement of U.** This may be read as **perpendicular.**

It is clear from the definition that if $x \in U^{\perp}$ then x is orthogonal to each element of U. We first show that $U^{\perp}$ is a subspace of V.

Since, $< \theta, u > = 0$ for all $u \in U$, we have, $\theta \in U^{\perp}$.

Let $v, w \in U^{\perp}$and $c \in F$. Then for any $u \in U$, we have,

$< cv + w, u > = c < v, u > \; + < w, u > = 0$ [$as < v, u > = < w, u > = 0$]

Therefore, $cv + w \in U^{\perp}$.

Hence, $U^{\perp}$ is a subspace of V.

Clearly, in$\mathbb{R}^2$, orthogonal complement of x- axis is y - axis as $< (a, 0), (0, b) > = a.0 + 0.b = 0$ and orthogonal complement of the line $y =$ is $y = -x$ as $< (a, a), (b, -b) >= ab - ab = 0$.

Similarly, in $\mathbb{R}^3$, orthogonal complement of x - axis is yz plane and vice versa as any point on $x -$ axis can be taken as $(a, 0,0)$ and any point on yz plane can be taken as $(0, b, c)$ and we have,

$$< (a, 0,0), (0, b, c) > = a.0 + 0.b + 0.c = 0$$

Example. In Euclidean space $\mathbb{R}^4$ with standard inner product, let Wbe the subspace generated by the vectors $(1,0,-1,-4)$and $(0,1,1,2)$. What is $W^{\perp}$?

Let $v_1 = (1,0,-1,-4)$, $v_2 = (0,1,1,2)$and $W = L\{v_1, v_2\}$.

Let $u = (a, b, c, d) \in W^{L}$. Then

$$< u, v_1 > = 0 => a - c - 4d = 0 \ldots (1)$$

And $$< u, v_2 > = 0 => b + c + 2d = 0 \ldots (2)$$

By (1)and (2), we have,

$(a, b, c, d) = (c + 4d, -c - 2d, c, d) = c(1,-1,1,0) + d(4,-2,0,1), c, d \in \mathbb{R}$

Hence, $W^{\perp} = span\,\{(1,-1,1,0), (4,-2,0,1)\}$.

We now study some basic properties of orthogonal complements

2.2.2 Theorem. If $S \subset T$ then $S^{\perp} \supset T^{\perp}$.

Proof. Let $x \in T^{\perp}$. Then $< x, t > = 0$ for all $\in T$.

Thus,$< x, s > = 0$ for all $s \in S$as $S \subset T$. Therefore, $x \in S^{\perp}$.

Hence, $S^{\perp} \supset T^{\perp}$.

2.2.3 Theorem. If S and T are subspaces, then $(S + T)^{\perp} = S^{\perp} \cap T^{\perp}$.

Proof. Since, $S \subset S + T$, we have, $(S + T)^{\perp} \subset S^{\perp}$. Similarly, $(S + T)^{\perp} \subset T^{\perp}$.

Therefore, $(S + T)^{\perp} \subset S^{\perp} \cap T^{\perp}$.

Let $x \in S^{\perp} \cap T^{\perp}$. Then $< x, s > = 0, \forall s \in Sand< x, t > = 0, \forall t \in T$.

Let $y \in S + T$. Then $y = s + t$ for some $s \in S$and $t \in T$.

Now, $< x, y > = < x, s + t > = < x, s > \; + < x, t > = 0 + 0 = 0$

So, $x \in (S + T)^{\perp}$ as y is arbitrary.

Therefore, $S^{\perp} \cap T^{\perp} \subset (S + T)^{\perp}$.

Hence, $(S + T)^{\perp} = S^{\perp} \cap T^{\perp}$ (proved).

It should be mentioned here that if U_1, U_2 are subspaces of V, then V is the direct sum of U_1 and U_2 (written $V = U_1 \oplus U_2$) if each element of V can be written in exactly one way as a vector in U_1 plus a vector in U_2. The next theorem shows that every subspace of an inner-product space leads to a natural direct sum decomposition of the whole space.

2.2.4 Theorem. Let U be a vector subspace of an inner product space V. **Then$V = U \oplus U^{\perp}$. That is, any $x \in V$ is of the form $x = u + u'$ with $u \in U$ and $u' \in U^{\perp}$ and this decomposition is unique.**

Proof. Let us consider an orthonormal basis of Uas $\{u_1, u_2, .., u_r\}$. Let $x \in V$. Let us define

$$u = < x, u_1 > u_1 + < x, u_2 > u_2 + \cdots + < x, u_r > u_r \in U$$

Let $u' = x - u$.

Then $< u', u_k > = < x - u, u_k > = < x, u_k > - < u, u_k >$

$$= < x, u_k > - < \sum_{i=1}^{k} < x, u_i > u_i , u_k >$$

$$= < x, u_k > - \sum_{i=1}^{k} < x, u_i > < u_i, u_k >$$

$$= < x, u_k > - < x, u_k >$$

$$(as < u_i, u_k > = 0 \; for \; i \neq k \; and < u_k, u_k > = 1)$$

$$= 0$$

Therefore, u' is orthogonal to each u_k for $k = 1,2,..,r$.

Let $w \in U$. Then $w = c_1u_1 + c_2u_2 + \cdots + c_ru_r$ where $c_i \in F$ for $i =$ 1,2,..,r. Now,

$$< u', w > = < u', c_1u_1 + c_2u_2 + \cdots + c_ru_r >$$

$$= c_1 < u', u_1 > + c_2 < u', u_2 > + \cdots + c_r < u', u_r >$$

$$= 0 \; (as < u', u_k > = 0 \; for \; k = 1,2,..,r)$$

Thus, u' is orthogonal to each element of U, in other words, $u' \in U^{\perp}$.

Hence, $x = u + u'$ where $u \in U$ and $u' \in U^{\perp}$.

To prove the uniqueness, let $x = u + u' = q + q'$ where $u, q \in U$ and $u', q' \in U^{\perp}$. Then

$$u - q = q' - u'$$

Since U and $U^{\perp}$ both are subspaces, we have, $u - q \in U$ and $q' - u' \in U^{\perp}$.

If, $z = u - q = q' - u'$ then $z \in U \cap U^{\perp}$. But $U \cap U^{\perp} = \{\theta\}$ as θ is the only vector which is orthogonal to itself. Hence, $z = \theta$, that is, $u = q, u' = q'$.

Thus the decomposition $x = u + u'$, $u \in U, u' \in U^{\perp}$ is unique.

Hence, $V = U \oplus U^{\perp}$ ∎

2.2.5 Corollary : If U is a subspace of an inner product space V, then $U = (U^{\perp})^{\perp}$.

Proof. Let $u \in U$. Then $< u, v > = 0$ for all $v \in U^{\perp}$. Thus u is orthogonal to all vectors of $U^{\perp}$. Therefore, $u \in (U^{\perp})^{\perp}$. Hence, $U \subseteq (U^{\perp})^{\perp}$.

Now, let $v \in (U^{\perp})^{\perp}$.

Then $v \in V$ and we can write $v = u + w$ where $u \in U, w \in U^{\perp}$.

But $u \in U \Rightarrow u \in (U^{\perp})^{\perp}$ (already proved). Therefore,

$v \in (U^{\perp})^{\perp}$, $u \in (U^{\perp})^{\perp}$ and $(U^{\perp})^{\perp}$ is a subspace, together imply $v - u \in (U^{\perp})^{\perp}$ which means that $v - u$ is orthogonal to itself. This implies $v - u = \theta$, in other words, $v = u \in U$. Thus, $(U^{\perp})^{\perp} \subseteq U$.

Hence, $U = (U^{\perp})^{\perp}$ ∎

Note: The expression $x = u + u'$ where $u \in U$ and $u' \in U^{\perp}$ is called the **orthogonal decomposition of the vector x with respect to U.** The inner product space V is called an **orthogonal direct sum** of U and $U^{\perp}$. What happens if we define a map $P_U : V \rightarrow U$ by $P_U(x) = u$? Let's come to the following definition.

2.2.6 Definition. Let U be a vector subspace of an inner product space V and $V = U \oplus U^{\perp}$. Then the **orthogonal projection P_U** of V onto U is the map $P_U(x) = u$ where $x = u + u'$ such that $u \in U$, and $u' \in U^{\perp}$.

In fact, in the proof of theorem 2.2.4, an expression of P_U in terms of an orthonormal basis of U has already been introduced. If $\{u_1, u_2, ..., u_r\}$ is an orthonormal basis of U, then

$$P_U(x) = u = < x, u_1 > u_1 + < x, u_2 > u_2 + \cdots + < x, u_r > u_r$$

Geometrically, the orthogonal projection of x onto U is the foot of the perpendicular drawn from x to U. What about $x - u$? $x - u = u' \in U^{\perp}$. Thus, $x - u$ is the orthogonal projection of x into $U^{\perp}$.

2.2.7 Theorem. Let W be a vector subspace of a finite dimensional inner product space V. Let $\{w_1, w_2, ..., w_r\}$ be an orthonormal basis of W. Let $\{u_1, u_2, ..., u_s\}$ be an orthonormal basis of $W^{\perp}$. Then $\{w_1, w_2, ..., w_r, u_1, u_2, ..., u_s\}$ is an orthonormal basis of V.

Solution. Since norm of each element of $\{w_1, w_2, ..., w_r, u_1, u_2, ..., u_s\}$ is 1 and

$$< w_i, w_j > = < u_i, u_j > = < w_i, u_j > = 0 \; for \; i \neq j$$

we see that the set $\{w_1, w_2, ..., w_r, u_1, u_2, ..., u_s\}$ is orthonormal and hence is linearly independent.

Let $v \in V$. Then $v = w + w'$ where $w \in W, w' \in W^{\perp}$. Since $\{w_1, w_2, ..., w_r\}$ and $\{u_1, u_2, ..., u_s\}$ are bases of W and $W^{\perp}$ respectively, there exist scalars $c_1, c_2, ..., c_r, d_1, d_2, ,, d_s$ such that

$$v = w + w' = c_1w_1 + c_2w_2 + \cdots + c_rw_r + d_1u_1 + d_2u_2 + \cdots + d_su_s$$

Thus, $v \in L\{w_1, w_2, ..., w_r, u_1, u_2, ..., u_s\}$.

Hence, $\{w_1, w_2, ..., w_r, u_1, u_2, ..., u_s\}$ is a basis of V ∎

2.2.8 Theorem. Let S be a subset of an inner product space V, and $W = L(S)$. Then $S^{\perp} = W^{\perp}$.

Proof. If $S = \emptyset$, then $W = \{\theta\}$ and hence, $S^{\perp} = V = W^{\perp}$. So, let $S \neq \emptyset$.

Now, $S \subseteq W \Rightarrow W^{\perp} \subseteq S^{\perp}$.

So, let $s' \in S^{\perp}$. Let $w \in W$.

Since, $L(S) = W$, there exist $s_1, s_2, \ldots, s_n \in S$ and scalars $c_1, c_2, \ldots, c_n$ such that $w = c_1 s_1 + c_2 s_2 + \cdots + c_n s_n$

Now, $< w, s' > = < c_1 s_1 + c_2 s_2 + \cdots + c_n s_n, s' >$

$$= c_1 < s_1, s' > + c_2 < s_2, s' > + \cdots + c_n < s_n, s' >$$

$$= 0 \; [< s_i, s' > = 0 \; for \; i = 1,2, \ldots, n \; as \; s' \in S^{\perp}]$$

Therefore, $s' \in W^{\perp}$. In other words, $S^{\perp} \subseteq W^{\perp}$.

Hence, $S^{\perp} = W^{\perp}$ ∎

Now, we may discuss a problem: given a subspace U of V and a point $v \in V$, find a point $u \in U$ such that $\|v - u\|$ is as small as possible. The next theorem shows that this minimization problem is solved by taking $u = P_U(v)$.

2.2.9 Theorem. Suppose U is a subspace of an inner product space V and $v \in V$. Then $\|v - P_U(v)\| \le \|v - u\|$, $\forall u \in U$

Furthermore, if $u \in U$ and the inequality above is an equality, then $u = P_U(v)$.

Proof. Let $v \in V$. Then $v = x + x'$ where $x \in U$ and $x' \in U^{\perp}$.

Thus, $P_U(v) = x$. Therefore, $v - P_U(v) = v - x = x' \in U^{\perp}$ and $P_U(v) - u \in U$ for all $u \in U$.

Hence, for any $u \in U$, we have, $< v - P_U(v), \; P_U(v) - u > = 0$. Therefore, by Pythagorean theorem,

$$\|v - P_U(v)\|^2 + \|P_U(v) - u\|^2 = \|v - P_U(v) + P_U(v) - u\|^2 = \|v - u\|^2$$

Thus, $\|v - u\|^2 \ge \|v - P_U(v)\|^2$.

Hence, $\|v - P_U(v)\| \le \|v - u\|$ for all $u \in U$.

Clearly, equality holds if and only if

$\|v - P_U(v)\|^2 + \|P_U(v) - u\|^2 = \|v - u\|^2$ which occurs if and only if $\|P_U(v) - u\| = 0$ if and only if $P_U(v) - u = \theta$ if and only if $P_U(v) = u$ ∎.

What is the geometrical interpretation of P_U? If $v \in V$ then $P_U(v) \in U$ is the unique element of U which is nearest to v as $\|v - P_U(v)\| \le \|v - u\|$ for all $u \in U$.

2.2.10 Bessel's Inequality : Let V be an inner product space, and let $S = \{u_1, u_2, \ldots, u_n\}$ be an orthonormal subset of V. Prove that for any $x \in V$,

$$\|x\|^2 \ge \sum_{i=1}^{n} |< x, u_i >|^2$$

Proof. Let $u = < x, u_1 > u_1 + < x, u_2 > u_2 + \cdots < x, u_n > u_n$

For any u_j, $j \in \{1,2, \ldots, n\}$,

we have,

$< x - u, u_j > = < x, u_j > -$

$$< (< x, u_1 > u_1 + < x, u_2 > u_2 + \cdots < x, u_n > u_n), u_j >$$

$$= < x, u_j > - < x, u_j > \; (as < u_i, u_j >$$

$$= 0 for \; i \neq j \; and < u_j, u_j > = 0$$

Thus, $u_j \in (x - u)^{\perp}$ for all $j = 1,2, \ldots, n$.

Since, $(x - u)^{\perp}$ is a subspace, we have,

$$u = < x, u_1 > u_1 + < x, u_2 > u_2 + \cdots < x, u_n > u_n \in (x - u)^{\perp}$$

Thus, $x = (x - u) + u$ where $< x - u, u > = 0$. So, by Pythagorean theorem,

$$\|x\|^2 = \|x - u\|^2 + \|u\|^2$$

which implies that $\|x\|^2 \ge \|u\|^2 (as \; \|x - u\|^2 \ge 0)$

i.e.

$$\|x\|^2 \ge \sum_{i=1}^{n} |< x, u_i >|^2$$

Hence the result.

Note : Bessel's inequality becomes an equality if and only $\|x - u\| = 0$, that is, if and only if

$$x = u \in L(S).$$

Solved Examples :

1. **Suppose U is a subspace of V. Prove that $U^{\perp} = \{\theta\}$ if and only if $U = V$.**

Solution. Since U is a subspace of V,

we have, $V = U \oplus U^{\perp}$.

Thus, $U^{\perp} = \{\theta\}$ if and only if $V = U$.

2. **Let $S = \{(1, 0, i), (1, 2, 1)\}$ in C^3. Compute $S^{\perp}$.**

Solution. Let $(a, b, c) \in S^{\perp}$.

Then $< (a, b, c), (1,0, i) > = 0 => a - ci = 0 \ldots..(1)$

And $< (a, b, c), (1,2,1) > = 0 => a + 2b + c = 0 \ldots.(ii)$

By (i), $a = ci$. Using it in (ii),

we get $b = -\frac{1}{2}(a + c) = -\frac{1}{2}c(1 + i)$.

Hence,

$$(a,b,c) = \left(ci, -\frac{c}{2}(1 + i), c\right) = c\left(i, -\frac{1}{2}(1 + i), 1\right) \quad c \in \mathbb{R}.$$

Hence, $S^{\perp} = span\left\{\left(i, -\frac{1}{2}(1 + i), 1\right)\right\}$.

3. **Find the orthogonal complement of the row space of the matrix**

$$A = \begin{pmatrix} 1 & 2 & 10 \\ 1 & 3 & 22 \\ 0 & 1 & 12 \end{pmatrix}.$$

Solution. Let us apply elementary row operations on A to bring it to row reduced form as follows :

$$A \to \begin{pmatrix} 1 & 2 & 10 \\ 0 & 1 & 12 \\ 0 & 1 & 12 \end{pmatrix} [R_2' = R_2 - R_1]$$

$$\to \begin{pmatrix} 1 & 0 & -1-4 \\ 0 & 1 & 1 \;\; 2 \\ 0 & 0 & 0 \;\; 0 \end{pmatrix} [R_1' = R_1 - 2R_2, R_3' = R_3 - 2R_2]$$

Let $v_1 = (1,0,-1,4)$ and $v_2 = (0,1,1,2)$. Then the row space of A is given by $W = span\,\{v_1, v_2\}$.

Let $u = (a,b,c,d) \in W^{\perp}$.

Then $< u, v_1 > = 0 => a - c - 4d = 0$ ... (1)

and$< u, v_2 > = 0 => b + c + 2d = 0$ ... (2)

By (1) and (2), we have,

$(a,b,c,d) = (c + 4d, -c - 2d, c, d) = c(1,-1,1,0) + d(4,-2,0,1), \quad c,d \in \mathbb{R}$

Therefore, $W^{\perp} = span\,\{(1,-1,1,0), (4,-2,0,1)\}$.

Hence the orthogonal complement of the row space of A is given by $span\,\{(1,-1,1,0), (4,-2,0,1)\}$.

4. **Let A be a real $m \times n$ matrix. Show that the solution space of the system of the equations $AX = 0$ is the orthogonal complement of the row space of A in the Euclidean space $\mathbb{R}^n$ with standard inner product.**

Solution. Let $A = (a_{ij})_{m \times n}$ be a real matrix and $R_1, R_2, \ldots, R_m$ be the row vectors of A where

$$R_i = (a_{i1}, a_{i2}, \ldots, a_{in}), \qquad i = 1,2, \ldots, m.$$

Then the row space of A is given by $R = L\{R_1, R_2, \ldots, R_m\}$.

Let W be the solution space of $AX = 0$, that is,

$$W = \{S \in \mathbb{R}^n : AS = 0\}.$$

Let $S = (s_1, s_2, \ldots, s_n) \in W$. Then$AS = 0$, that is ,

$$a_{i1}s_1 + a_{i2}s_2 + \cdots + a_{in}s_n = 0, i = 1,2, \ldots, m.$$

Thus S is orthogonal to each R_i for $i = 1,2, \ldots, m$. Hence, $S \in R^{\perp}$.

Therefore, $W \subseteq R^{\perp}$.

Conversely, let $U = (u_1, u_2, \ldots, u_n) \in R^{\perp}$. Then U is orthogonal to each R_i for $i = 1,2, \ldots, m$.

Thus, $a_{i1}u_1 + a_{i2}u_2 + \cdots + a_{in}u_n = 0, \; i = 1,2, \ldots, m$.

Therefore, $U \in W$. In other words, $R^{\perp} \subseteq W$.

Hence $W = R^{\perp}$, that is, the solution space of the system $AX = 0$ is the orthogonal complement of the row space of A.

5. **Let β be a basis for a subspace W of a finite dimensional inner product space V, and let $z \in V$. Prove that$z \in W^{\perp}$if and only if $< z, v > = 0$ for every $v \in \beta$.**

Solution. Since β is a basis for$W$, so, $v \in \beta, z \in W^{\perp}$implies $< z, v > = 0$. Thus the condition is necessary.

Conversely, let $< z, v > = 0$ for all $v \in \beta$. We shall show that $z \in W^{\perp}$.

Let $\beta = \{v_1, v_2, \ldots, v_k\}$. Let $w \in W$. Then there exist scalars $c_1, c_2, \ldots, c_k$ such that

$$w = c_1v_1 + c_2v_2 + \cdots + c_kv_k$$

Now,

$$< z, w > = < z, c_1v_1 + c_2v_2 + \cdots + c_kv_k >$$
$$= \bar{c_1} < z, v_1 > + \bar{c_2} < z, v_2 > + \cdots + \bar{c_k} < z, v_k >$$
$$= 0 \;\; [\,as < z, v > = 0 \;\forall v \in \beta].$$

Hence, $z \in W^{\perp}$.

6. **Find the orthogonal projection of the vector $u = (2,6)$ on the subspace $W = \{(x,y) : y = 4x\}$ of the vector space $\mathbb{R}^2$ with standard inner product.**

Solution. Here, $W = \{\,x(1,4) : x \in \mathbb{R}\} = L\{(1,4)\}$. Thus, basis for W is given by $\{(1,4)\}$. Therefore, an orthonormal basis for W is given by $e = \frac{1}{\sqrt{17}}(1,4)$.

Hence, the projection of u on W is given by

$$< u,e > e = < (2,6), \frac{1}{\sqrt{17}}(1,4) > = \frac{1}{\sqrt{17}} < (2,6),(1,4) > \frac{1}{\sqrt{17}}(1,4)$$
$$= \frac{26}{17}(1,4).$$

Exercise

1. Find the orthogonal complement of the row space of the matrix
$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 3 & 5 \\ 3 & 4 & 7 \end{pmatrix}.$$
2. Let $W = span\,\{(i,0,1)\}$in $\mathbb{C}^3$. Find orthonormal bases for W and $W^{\perp}$.
3. Let W_1 and W_2 be subspaces of a finite- dimensional inner product space. Prove that $(W_1 + W_2)^{\perp} = W_1^{\perp} \cap W_2^{\perp}$ and $(W_1 \cap W_2)^{\perp} = W_1^{\perp} + W_2^{\perp}$.
4. Find the orthogonal projection of the vector $u = (2,1,3)$ on the subspace$W = \{(x,y,z) : x + 3y - 2z = 0\}$ of the vector space $\mathbb{R}^3$ with standard inner product.

2.3 ADJOINT OF A LINEAR OPERATOR AND ITS BASIC PROPERTIES

We know that if V be a vector space over a field F then a mapping $T : V \to V$ is called a **Linear operator on V** if for all $x,y \in V$ and for all $c \in F$

(i) $T(x+y) = T(x) + T(y)$ and

(ii) $T(cx) = cT(x)$

or, equivalently $T(cx+y) = cT(x) + T(y)$.

Properties of linear operators are discussed in detail in my book *Ring Theory and Linear Algebra*. Readers are advised to go through that book if they need any support. Here we study some advanced part of it. In addition, we must take into account the *linear functional* which is defined as a linear mapping $f : V \to F$ where V is a vector space over a field F. For example, for some fixed $y \in V$ the mapping $f : V \to F$ by $f(x) = < x,y >$ is a linear functional. It is very interesting that if V is finite dimensional then every linear functional from V into F is of this form. Theorem 2.3.4 will support it.

Let's start with the following definition.

2.3.1 Definition. Let T be a linear operator on an inner product space V. Then we say that **T has an adjoint on V** if there exists a linear operator T^* on V such that $< T(x),y > = < x,T^*(y) >$for all x and y in V.

One may doubt the existence of T^* for a given linear operator T on V. Following result will help us.

2.3.4 Theorem. Let V be a finite dimensional inner product space over a field F and let $g : V \to F$ be a linear functional. Then there exists a unique vector $y \in V$ such that $g(x) = < x,y >$ for all $x \in V$.

Proof. Let $B = \{v_1, v_2, \dots, v_n\}$ be an orthonormal basis of V, and let

$$y = \sum_{i=1}^{n} \overline{g(v_i)}\, v_i$$

Let us define $h : V \to F$ by $h(x) = < x,y >$. Clearly, h is linear. Now,

$$h(v_j) = < v_j, y > = < v_j, \sum_{i=1}^{n} \overline{g(v_i)}\, v_i > = \sum_{i=1}^{n} g(v_i) < v_j, v_i >$$
$$= g(v_1) < v_j, v_1 > + g(v_2) < v_j, v_2 > + \dots + g(v_j) < v_j, v_j > + \dots + g(v_n) < v_j, v_n >$$
$$= g(v_j)[\,as < v_j, v_i > = 0\ for\ i \neq j\ and < v_j, v_j > = 1]$$

So, $h(v_j) = g(v_j)$ for $j = 1,2,\dots,n$ and hence, $h(x) = g(x)$ for all $x \in V$. Thus, $h = g$.

To show the uniqueness of y, let $g(x) = < x,y' >$ for some $y' \in V$ and for all $x \in V$. Then for $x \in V$,

$$g(x) = < x,y > = < x,y' >$$

Since this is true for all $x \in V$, we have, $y = y'$.

Hence proved.

2.3.5 Example : Let $V = \mathbb{R}^3$, define $g : V \to \mathbb{R}$ by $g(x,y,z) = x - 2y + 4z$. We want to find a vector $v \in V$ such that $g(u) = < u,v >$ for all $u \in V$.

We know, $\{e_1, e_2, e_3\}$ is an orthonormal basis of $\mathbb{R}^3$ where

$$e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1).$$

So, $g(e_1) = 1$, $g(e_2) = -2$, $g(e_3) = 4$. We take

$$v = g(e_1)e_1 + g(e_2)e_2 + g(e_3)e_3$$
$$= 1(1,0,0) + (-2)(0,1,0) + 4(0,0,1) = (1,-2,4)$$

Thus, for any $u = (x,y,z) \in V$, we have

$$g(u) = < u, v > = < (x,y,z),(1,-2,4) > = x - 2y + 4z \blacksquare$$

Let's come to the important theorem.

2.3.6 Theorem. Let V be a finite dimensional inner product space and let T be a linear operator on V. Then there exists a unique function $T^* : V \to V$ such that $< T(x), y > = < x, T^*(y) >$ for all $x, y \in V$. Furthermore, T^* is linear.

Proof. Let $y \in V$. Define $g : V \to F$ by $g(x) = < T(x), y >$ for all $x \in V$. We first show that g is linear. Let $x_1, x_2 \in V$ and $c \in F$. Then

$$\begin{aligned} g(cx_1 + x_2) &= < T(cx_1 + x_2), y > \\ &= < cT(x_1) + T(x_2), y > [\text{ as } T \text{ is linear}] \\ &= c < T(x_1), y > + < T(x_2), y > \\ &= cg(x_1) + g(x_2) \end{aligned}$$

Hence, g is linear.

Therefore, by theorem 2.3.4 there exists a unique vector y' in V such that $g(x) = < x, y' >$. In other words, $g(x) = < T(x), y > = < x, y' >$.

Let us define $T^* : V \to V$ by $T^*(y) = y'$, that is, $< T(x), y > = < x, T^*(y) >$

Now, we show that T^* is linear.

Let $y_1, y_2 \in V$ and $c \in F$. Then for $x \in V$, we have,

$$\begin{aligned} & < x, T^*(cy_1 + y_2) > \\ &= < T(x), cy_1 + y_2 > \\ &= \bar{c} < T(x), y_1 > + < T(x), y_2 > \\ &= \bar{c} < x, T^*(y_1) > + < x, T^*(y_2) > \\ &= < x, cT^*(y_1) > + < x, T^*(y_2) > \\ &= < x, cT^*(y_1) + T^*(y_2) > \end{aligned}$$

Since, x is arbitrary, we have, $T^*(cy_1 + y_2) = cT^*(y_1) + T^*(y_2)$. Therefore, T^* is linear.

To prove the uniqueness of T^*, let $S : V \to V$ be such that for all $x, y \in V$

$$< T(x), y > = < x, S(y) >$$

Thus, $< x, S(y) > = < x, T^*(y) >$ for all $x, y \in V$ which shows that $S = T^*$.

Hence the theorem ■

The linear operator T^* used in the above theorem is the **adjoint**of the operator T.

Basic properties of the adjoin of a linear operator

2.3.7 Theorem. Let V be a finite dimensional inner product space over a field ; S and T be two linear operators on V, and $a \in F$. Then

(i) $(T^*)^* = T$

(ii) $(S+T)^* = S^* + T^*$

(iii) $(aT^*)^* = \bar{a}T^*$

(iv) $(ST)^* = T^*S^*$

Proof:

(i) Let $x, y \in V$.

Then $< T^*(y), x > = < y, T^{**}(x) >$

$$\begin{aligned} &\Rightarrow \overline{< T^*(y), x >} = \overline{< y, T^{**}(x) >} \\ &\Rightarrow < x, T^*(y) > = < T^{**}(x), y > \\ &\Rightarrow < T(x), y > = < T^{**}(x), y > \end{aligned}$$

Hence, $T(x) = T^{**}(x)$, for all $x \in V$.

Therefore, $T = T^{**}$.

(ii) Let $x, y \in V$.

Now, $< x, (S+T)^*(y) > = < (S+T)(x), y >$

$$\begin{aligned} &= < S(x) + T(x), y > \\ &= < S(x), y > + < T(x), y > \\ &= < x, S^*(y) > + < x, T^*(y) > \\ &= < x, S^*(y) + T^*(y) > \\ &= < x, (S^* + T^*)(y) > \end{aligned}$$

Hence, $(S+T)^*(y) = (S^* + T^*)(y)$ for all $y \in V$.

Therefore, $(S+T)^* = S^* + T^*$.

(iii) For $x, y \in V$,

we have, $< x, (aT)^*(y) > = < aT(x), y >$

$$\begin{aligned} &= a < T(x), y > \\ &= a < x.T^*(y) > = < x, \bar{a}T^*(y) > \end{aligned}$$

So, $(aT)^*(y) = \bar{a}T^*(y)$ for all $y \in V$.

Hence, $(aT)^* = \bar{a}T^*$.

(iv) For $x, y \in V$,
we have, $< x, (ST)^*(y) > = < ST(x), y >$
$= < S(T(x)), y >$
$= < T(x), S^*(y) >$
$= < x, (T^*S^*)(y) >$

Thus, $(ST)^*(y) = T^*S^*(y)$ for all $y \in V$.

Hence, $(ST)^* = T^*S^*$ ■

The next result shows the relationship between the null space and therangeofa linear operator anditsadjoint. But before that I think it is better to recap the definitions of null space and range of a linear operator.

If $T : V \to V$ be a linear operator where V is a vector space over a field F then **null space of T or kernel of T**, denoted by *null T* or *kerT*, is defined as

$$null\ T = \ker T = \{v \in V : T(v) = \theta\}$$

and $T(V) = \{w \in V : \exists\ v \in V\ such\ that\ T(v) = w\}$ is known as **range of T.**

Thesymbol$<=>$ means"ifand only if"; this symbol could also be read to mean "is equivalent to".

2.3.8 Theorem. Let $T : V \to V$ be a linear operator, V being an inner product space over a field F, then

(i) $null\ T^* = (rangeT)^{\perp}$

(ii) $range\ T^* = (null\ T)^{\perp}$

(ii) $null\ T = (range\ T^*)^{\perp}$

(iii) $range\ T = (null\ T^*)^{\perp}$

Proof.

(i) Let $w \in V$.

Then $w \in null\ T^*$

$$<=> \quad T^*(w) = \theta$$
$$<=> < v, T^*(w) > = 0\ \forall v \in V$$
$$<=> < T(v), w > = 0, \forall v \in V$$
$$<=> \quad w \in (rangeT)^{\perp}$$

Hence, $null\ T^* = (range\ T)^{\perp}$.

(ii) If we replace T by T^* in (i), and use $(T^*)^* = T$, we get,

$$null\ T = (range\ T^*)^{\perp}$$

Taking orthogonal complement on both sides, we get

$$range\ T^* = (null\ T)^{\perp}[\ as\ (A^{\perp})^{\perp} = A]$$

(iii) already proved in (ii)

(iv) By (i), $null\ T^* = (range\ T)^{\perp}$. Taking orthogonal complements on both sides

$range\ T = (null\ T^*)^{\perp}$.

2.3.9 Definition. A linear operator T defined on an inner product space Vover a field F is called **self adjoint (orHermitian)** if $T^* = T$.

2.3.10 Theorem. Let V be a finite dimensional inner product space over a field F and S, T be two linear operators on V then

(i) If T is invertible then T^* is invertible and $(T^*)^{-1} = (T^{-1})^*$

(ii) If S and T are self adjoint then $S + T$ is also self adjoint

(iii) If S and$T$ are self adjoint then ST is self adjoint if and only if $ST = TS$

(iv) For any $a \in F$ and for any self adjoint operator T, aT is self adjoint if and only if a is real.

Proof.

(i) Since, T is invertible, we have, $TT^{-1} = I = T^{-1}T$. Then by theorem 2.3.7,

$$(TT^{-1})^* = I^* = I => (T^{-1})^*T^* = I$$

Similarly, $T^*(T^{-1})^* = I$.

Hence, T^* is invertible and $(T^*)^{-1} = (T^{-1})^*$.

(ii) Since S, T are self adjoint operators, we have, $S^* = S$ and $T^* = T$.

By theorem 2.3.7, $(S + T)^* = S^* + T^* = S + T$.

Hence, $S + T$ is self adjoint.

(iii) We have, $S^* = S, T^* = T$. By theorem 2.3.7

$$(ST)^* = ST <=> T^*S^* = ST <=> TS = ST.$$

Hence, ST is self adjoint if and only if $ST = TS$.

(iv) Let $a \in F$ and T be a self adjoint operator i.e. $T^* = T$. Now, by theorem 2.3.7,

$$(aT)^* = aT <=> \bar{a}T^* = aT <=> \bar{a}\,T = aT <=> \bar{a} = a <=> a\ is\ real$$

Hence proved.

Solved example :

1. **Let V be an inner product space over the field F. Fix a vector $v \in V$. Define a linear functional $T : V \to F$ by $T(u) = < u, v >$. For $a \in F$, find a formula for $T^*(a)$.**

Solution. It is clear that T^* is a mapping from F to V. Thus, for a fixed $a \in F$, we see that $T^*(a)$ is a unique vector in V such that

$< T(u), a > = < u, T^*(a) >$ for all $u \in V$.....(1)

The inner product on the right is the inner product defined in V, but the inner product on the left is the usual inner product defined on F; the product of the entry in the first slot with the complex conjugate of the entry in the second slot. Thus,

$$< T(u), a > = [T(u)]\bar{a} = < u, v > \bar{a} = < u, av > \cdots (2).$$

Using (1) and (2), we can say that

$$< u, T^*(a) > = < u, av >$$

for all $u \in V$. Hence, we have, $T^*(a) = av$ ■

2. **Let n be a fixed positive integer. Define $T \in \mathcal{L}(F^n)$ by $T(z_1, z_2, \ldots, z_n) = (0, z_1, z_2, \ldots, z_{n-1})$ find a formula for $T^*(z_1, z_2, \ldots, z_n)$.**

Solution. Let us fix $(z_1, z_2, \ldots, z_n)$. Then for each $(w_1, w_2, \ldots, w_n)$, we have,

$$\begin{aligned} &< (w_1, w_2, \ldots, w_n), T^*(z_1, z_2, \ldots, z_n) > \\ &= < T(w_1, w_2, \ldots, w_n), (z_1, z_2, \ldots, z_n) > \\ &= < (0, w_1, w_2, \ldots, w_{n-1}), (z_1, z_2, \ldots, z_n) > \\ &= w_1\overline{z_2} + w_2\overline{z_3} + \cdots + w_{n-1}\overline{z_n} + w_n.0 \\ &= < (w_1, w_2, \ldots, w_n), (z_2, z_3, \ldots, z_n, 0) > \end{aligned}$$

Hence, we have, $T^*(z_1, z_2, \ldots, z_n) = (z_2, z_3, \ldots, z_n, 0)$ ■

3. **Suppose $T \in \mathcal{L}(V)$ and $\lambda \in F$. Prove that λ is an eigenvalue of T if and only if $\bar{\lambda}$ is an eigenvalue of T^*.**

Solution. We have, λ is an eigenvalue of T

$\Leftrightarrow T - \lambda I$ is invertible $\Leftrightarrow$ there exists $S \in \mathcal{L}(V)$,

such that $S(T - \lambda I) = (T - \lambda I)S = I \Leftrightarrow (T - \lambda I)^* S^* = S^*(T - \lambda I)^* = I$

[as $(UV)^* = V^*U^*, \forall\, U, V \in \mathcal{L}(V)$]

$\Leftrightarrow (T - \lambda I)^*$ is invertible

$\Leftrightarrow T^* - \bar{\lambda} I$ is invertible

$\Leftrightarrow \bar{\lambda}$ is an eigenvalue of T^*.

Hence, the result.

4. **Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V. Prove that U is invariant under T if and only if $U^\perp$ is invariant under T^*.**

Solution. First suppose that U is invariant under T, that is, for $u \in U$, we have, $T(u) \in U$.

Let $v \in U^\perp$. Thus, $< u, v > = 0$ for all $u \in U$.

We shall show that $T^*(v) \in U^\perp$, in other words, $< u, T^*(v) > = 0$ for all $u \in U$. Now, for $u \in U$,

$$< u, T^*(v) > = < T(u), v > = 0\ [\, as\ T(u) \in U].$$

Hence, $U^\perp$ is invariant under T^*.

Conversely, let $U^\perp$ is invariant under T^*. Then by the first part, $(U^\perp)^\perp$ is invariant under $(T^*)^*$, that is, U is invariant under T as $(U^\perp)^\perp = U$ and $(T^*)^* = T$ ■

5. **Suppose $T \in \mathcal{L}(V, W)$. Prove that (a) T is injective if and only if T^* is surjective and (b) T is surjective if and only if T^* is injective.**

Solution. (a) Clearly,

T is injective $\Leftrightarrow$ null $T = \{0\}$

$\Leftrightarrow$ (range $T^*)^\perp = \{0\}$

$\Leftrightarrow$ range $T^* = W$

$\Leftrightarrow T^*$ is surjective.

(b) In (a) replace T by T^* and use the fact $(T^*)^* = T$.

6. **Prove that**

$$\dim null\, T^* = \dim null\, T + \dim W - \dim V$$

and

$$\dim range\, T^* = \dim range\, T$$

for every $T \in \mathcal{L}(V, W)$.

Solution. Let $T \in \mathcal{L}(V, W)$.

Then

$\dim null\, T^* = \dim(range\ T)^\perp$

$= \dim W - \dim(range\ T)$

[as $\dim(range\ T) + \dim(range T)^\perp = \dim W$]

$= \dim null\, T + \dim W - \dim V$

by Sylvester's law, $\dim null\, T + \dim range\, T = \dim V$]

To prove the last part, we have

$$\dim(range\, T^*) = \dim W - \dim(null\, T^*)$$
$$= \dim V - \dim(null\, T)[\text{ by 1st part}]$$
$$= \dim(range\, T)$$

Hence proved.

Exercise

1. Let V be the space $\mathbb{C}^2$, with standard inner product. Let T be the linear operator defined by

   $Te_1 = (1,-2)$, $Te_2 = (i,-1)$. If $v = (x_1, x_2)$, find T^*v.

2. Let T be a linear operator on $\mathbb{C}^2$ defined by $Te_1 = (1+i, 2)$, $Te_2 = (i, i)$. Using the standard inner product, find the matrix of T^* in the standard ordered basis. Does T commute with T^*?

3. Let V be a finite dimensional inner product space and T a linear operator on V. If T is invertible, show that T^* is invertible and $(T^*)^{-1} = (T^{-1})^*$.

4. Let V be an inner product space and β, γ fixed vectors in V. Show that $T(\alpha) = <\alpha, \beta> \gamma$ defines a linear operator on V. Show that T has an adjoint, and describe T^* explicitly.

5. Show that the product of two self-adjoint operators is self-adjoint if and only if the two operators commute.

6. Let V be a finite dimensional complex inner product space, and let T be a linear operator on V. Prove that T is self-adjoint if and only if $<Tv, v>$ is real for each $v \in V$.

2.4 BILINEAR AND QUADRATIC FORMS

So far, we have studied linear functionals whose domain is V, a vector space over a field F, and codomain is F satisfying linear properties. Now, we shall study those functions which are defined on $V \times V$ and have codomain as F satisfying some properties described below.

2.4.1 Definition. Let V be a vector space over a field F. A function $f: V \times V \to F$ is called a **bilinear form on** V if

(i) $f(cx_1 + x_2, y) = cf(x_1, y) + f(x_2, y)$ for all $x_1, x_2, y \in V$ and $c \in F$

(ii) $f(x, dy_1 + y_2) = d\, f(x, y_1) + f(x, y_2)$ for all $x, y_1, y_2 \in V$ and $d \in F$.

Thus, f is bilinear on V if f is linear in each variable when the other variable is held fixed. We denote the set of all bilinear forms on V by $\mathcal{B}(V)$.

2.4.2 Examples :

1. If V be a real inner product space, that is, a Euclidean space, then $f: V \times V \to \mathbb{R}$ defined by $f(x, y) = <x, y>$ is a bilinear functional.

Clearly, for $x_1, x_2, y \in V$ and $c \in \mathbb{R}$, we have,

$$f(cx_1 + x_2, y) = <cx_1 + x_2, y> = c<x_1, y> + <x_2, y>$$
$$= cf(x_1, y) + f(x_2, y)$$

and for $x, y_1, y_2 \in V$ and $d \in \mathbb{R}$, we have,

$$f(x, dy_1 + y_2) = <x, dy_1 + y_2>$$
$$= d<x, y_1> + <x, y_2>$$
$$= df(x, y_1) + f(x, y_2)$$

Therefore, f is bilinear.

Note : It is clear that f, as defined above is not bilinear if the underlying field F is complex as in that case $\bar{d}$ will come instead of d.

2. Let $V = F^n$, where the vectors are considered as column vectors. For any $A \in M_{n\times n}(F)$[that is, for an $n \times n$ matrix A whose entries are from F], define $f: V \times V \to F$ by

$$f(x, y) = x^t Ay, \quad for\ x, y \in V.$$

Look, here x and y are $n \times 1$ matrices and A is an $n \times n$ matrix. Thus, $f(x, y) = x^t Ay$ is a 1×1 matrix, that is, a scalar, a member of F.

Now, for $x_1, x_2, y \in V$ and $c \in F$, we have,

$$f(ax_1 + x_2, y) = (ax_1 + x_2)^t Ay = (ax_1^t + x_2^t)Ay$$
$$= ax_1^t Ay + x_2^t Ay = af(x_1, y) + f(x_2, y)$$

Again, for $x, y_1, y_2 \in V$ and for $d \in F$, we have,

$$f(x, dy_1 + y_2) = x^t A(dy_1 + y_2)$$
$$= d\, x^t Ay_1 + x^t Ay_2$$
$$= df(x, y_1) + f(x, y_2)$$

Hence, f is a bilinear form on V ■

2.4.3 Theorem. Let f be a bilinear form on a vector space V over a field F. Then

(i) If, for any $x \in V$, the functions $L_x, R_x : V \to F$ are defined by

$$L_x(y) = f(x,y), \text{ and } R_x(y) = f(y,x) \ \forall y \in V$$

then L_x and R_x are linear.

(ii) $f(\theta, x) = f(x, \theta) = 0$ for all $x \in V$.

(iii) For all $x, y, z, w \in V$,

$$f(x+y, z+w) = f(x,z) + f(x,w) + f(y,z) + f(y,w)$$

(iv) If $g : V \times V \to F$ is defined by $g(x,y) = f(y,x)$, then g is a bilinear form.

Proof.

(i) For $y_1, y_2 \in V$ and for $d \in F$, we have,

$$L_x(dy_1 + y_2) = f(x, dy_1 + y_2)$$
$$= d\, f(x, y_1) + f(x, y_2) \; [\text{ as } f \text{ is bilinear }]$$
$$= d\, L_x(y_1) + L_x(y_2)$$

Therefore, L_x is linear. Similarly, it can be proved that R_x is linear.

(ii) For any $v \in V$, we have, $\theta = 0.v$ and hence , for any $x \in V$, we have,

$$f(\theta, x) = f(0.v, x) = 0 f(v, x) = 0$$

Similarly, $f(x, \theta) = 0$.

(iii) For $x, y, z, w \in F$, we have,

$$f(x+y, z+w) = f(x, z+w) + f(y, z+w)$$
$$= f(x,z) + f(x,w) + f(y,z) + f(y,w)$$

(iv) Since f is bilinear on V, for $x_1, x_2, y \in V$ and $c \in F$, we have,

$$g(cx_1 + x_2, y) = f(y, cx_1 + x_2)$$
$$= cf(y, x_1) + f(y, x_2)$$
$$= c\, g(x_1 y) + g(x_2, y)$$

Similarly, for $x, y_1, y_2 \in V$, $d \in F$, we have,

$$g(x, dy_1 + y_2) = f(dy_1 + y_2, x) = df(y_1, x) + f(y_2, x)$$
$$= dg(x, y_1) + g(x, y_2)$$

Hence, g is a bilinear form on V.

2.4.4 Matrix representation of a bilinear form

Let $\beta = \{v_1, v_2, \dots, v_n\}$ be an ordered basis of a vector space V over a field F and f be a bilinear form defined on V. Then, we can associate with f an $n \times n$ matrix $A = (a_{ij})_{n \times n}$ where a_{ij} is defined by

$$a_{ij} = f(v_i, v_j), \qquad i,j = 1,2,..,n$$

The matrix A above is called the **matrix representation** of f with respect to the ordered basis β and is denoted by $\psi_\beta(f)$.

Example. Let $x, y \in \mathbb{R}^2$ where $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$.

Define $f : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) = < x, y > = < \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} >$

$$= x_1 y_1 + x_2 y_2$$

Let $B = \{e_1, e_2\}$ be standard ordered basis of $\mathbb{R}^2$ where $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

We wish to associate a matrix $A = (a_{ij})_{2\times 2}$ matrix with f with respect to B,

where

$$a_{11} = < \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} > = 1.1. + 0.0 = 1$$
$$a_{12} = < \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} > = 0$$
$$a_{21} = < \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} > = 0$$
$$a_{22} = < \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} > = 1$$

Hence, the matrix representation of f with respect to the ordered basis B is given by

$$\psi_B(f) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \blacksquare$$

2.5 SYMMETRIC BILINEAR FORMS

2.5.1 Definition. A bilinear form f on a vector space V is a **symmetric bilinear** form if $f(x,y) = f(y,x)$ for all $x, y \in V$.

As the name suggests, *matrix associated with a symmetric bilinear form is symmetric*, follows from the following theorem.

2.5.2 Theorem. Let f be a bilinear form on a finite dimensional vector space V and let B be an ordered basis for V. If f is symmetric then $\psi_B(f)$ is symmetric.

Proof. Let $B = \{v_1, v_2, .., v_n\}$ be an ordered basis for V and $A = \psi_B(f)$.

Thus, $A = (a_{ij})_{n\times n}$ where $a_{ij} = f(v_i, v_j)$.

If f is symmetric then $f(x,y) = f(y,x)$ for all $x, y \in V$. Therefore,

$$a_{ij} = f(v_i, v_j) = f(v_j, v_i) = a_{ji}, \qquad i,j \in \{1,2,\dots,n\}$$

Hence, A is symmetric■

2.5.3 Definition. A bilinear form f on a finite dimensional vector space V is called diagonalisableif there is an ordered basis B for V such that $\psi_B(f)$ is a diagonal matrix.

2.5.4 Definition. Let $A, B \in M_{n\times n}(F)$. Then B is said to be **congruent** to A if there exists an invertible matrix $Q \in M_{n\times n}(F)$ such that $B = Q^t AQ$.

The question is *how can you relate congruence to the matrix representation of a bilinear form?* Answer lies in the next theorem.

2.5.5 Theorem. Let V be a finite dimensional vector space with ordered bases $\beta = \{v_1, v_2, \dots, v_n\}$ and $\gamma = \{w_1, w_2, \dots, w_n\}$ and let Q be the change of coordinate matrix changing γ −coordinates into β −coordinates. Then for any bilinear form $f$on V, the matrix $\psi_\gamma(f)$ is congruent to $\psi_\beta(f)$.

Proof. Let $A = \psi_\beta(f)$and $B = \psi_\gamma(f)$. Since, $w_i \in V$ and β is an ordered basis for V, there exists scalars $Q_{1i}, Q_{2i}, \dots, Q_{ni}$ such that

$$w_i = Q_{1i}v_1 + Q_{2i}v_2 + \dots + Q_{ni}\, v_i = \sum_{k=1}^{n} Q_{ki}\, v_k$$

Similarly, we have,

$$w_j = \sum_{r=1}^{n} Q_{rj}\, v_r$$

Then, $B_{ij} = f(w_i, w_j) = f(\sum_{k=1}^{n} Q_{ki} v_k, w_j)$

$$= \sum_{k=1}^{n} Q_{ki} f(v_k, w_j)$$

$$= \sum_{k=1}^{n} Q_{ki} f(v_k, \sum_{r=1}^{n} Q_{rj}\, v_r)$$

$$= \sum_{k=1}^{n} Q_{ki} \sum_{r=1}^{n} Q_{rj} f(v_k, v_r)$$

$$= \sum_{k=1}^{n} Q_{ki} \sum_{r=1}^{n} Q_{rj}\, A_{kr}$$

$$= \sum_{k=1}^{n} Q_{ki} \sum_{r=1}^{n} A_{kr}\, Q_{rj}$$

$$= \sum_{k=1}^{n} Q_{ki}\, (AQ)_{kj}$$

$$= \sum_{k=1}^{n} Q^t_{ik}\, (AQ)_{kj}$$

$$= (Q^t AQ)_{ij}$$

Therefore, $B = Q^t AQ$

Hence, B is congruent to A, that is, $\psi_\gamma(f)$ is congruent to $\psi_\beta(f)$.

Now, we shall state a theorem without proof (beyond the scope of the syllabus)

2.5.6 Theorem. Let V be a finite dimensional vector space over a field not of characteristic 2 (i.e. $2.1 = 2 \neq 0$). Then a bilinear form on V is diagonalisable if and only if it is symmetric. Moreover, if $A \in M_{n\times n}(F)$ is a symmetric matrix then A is congruent to a diagonal matrix.

2.6 DIAGONALIZATION OF SYMMETRIC MATRICES

Let A be a symmetric $n \times n$ matrix with entries from a field F, not of characteristic 2. Ok, my dear friends, let us come to an agreement, *unless otherwise stated we take F as $\mathbb{R}$, the field of real numbers.*by the last part of theorem 2.5.6, we can say that A is congruent to some diagonal $n \times n$ matrix, say D, that is, there exists an invertible $n \times n$ matrix Q, such that $Q^t AQ = D$. But how to find $Q$and D? There is a simple method described below.

Use a sequence of elementary column operations and corresponding row operations to change the $n \times 2n$ matrix $(A|I)$into the form $(D|B)$, where D is a diagonal matrix and $B = Q^t$. Then we have, $D = Q^t AQ$. I think an example will clarify it.

Let

$$A = \begin{pmatrix} 1 & -1 & 3 \\ -1 & 2 & 1 \\ 3 & 1 & 1 \end{pmatrix}$$

Here A is a 3×3 real symmetric matrix. We start with a 3×6 matrix $(A|I)$ as follows:

$$(A|I) = \left(\begin{array}{ccc|ccc} 1 & -1 & 3 & 1 & 0 & 0 \\ -1 & 2 & 1 & 0 & 1 & 0 \\ 3 & 1 & 1 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{C_2 + C_1} \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 1 & 0 & 0 \\ -1 & 1 & 1 & 0 & 1 & 0 \\ 3 & 4 & 1 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_2 + R_1} \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 1 & 0 \\ 3 & 4 & 1 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{C_3 - 3C_1} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 1 & 0 \\ 3 & 4 & -8 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_3 - 3R_1} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 1 & 0 \\ 0 & 4 & -8 & -3 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{C_3 - 4C_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 4 & -24 & -3 & 0 & 1 \end{array}\right) \xrightarrow{R_3 - 4R_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & -24 & -7 & -4 & 1 \end{array}\right)$$

$$= (D|Q^t)$$

Where

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -24 \end{pmatrix}, \; Q^t = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -7 & -4 & 1 \end{pmatrix} \text{and therefore,} Q = \begin{pmatrix} 1 & 1 & -7 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{pmatrix}$$

Hence, $Q^t A Q = D$.

Please, notice that right part of the 3×6 matrix changes only when row operations are done.

2.7 QUADRATIC FORMS

Quadratic forms are homogeneous polynomials of degree 2 in several variables like

$2x^2 + y^2 - 3z^2 + 2xy + yz$.They occur in the study of conics in geometry, energy in Physics, and have wide applications in various subjects including Statistics. But here we try to define a quadratic form with the help of a bilinear form.

2.7.1 Definition. Let V be a vector space over $\mathbb{R}$. A function $q : V \to \mathbb{R}$ is called a **quadratic form** if there exists a symmetric bilinear form f on V (i.e., $f \in \mathcal{B}(V)$) such that

$$q(x) = f(x,x) \text{ for all } x \in V \; \blacksquare$$

If q is given, how can we find f? Let's try.

$$q(x+y) = f(x+y, x+y)$$
$$= f(x, x+y) + f(y, x+y)$$
$$= f(x,x) + f(x,y) + f(y,x) + f(y,y)$$
$$= q(x) + 2f(x,y) + q(y)[\, as\ f(x,y) = f(y,x)]$$

Hence, $f(x,y) = \frac{1}{2}[\,q(x+y) - q(x) - q(y)]$

2.7.2 Examples

The classic example of a quadratic form is the homogeneous second degree polynomial of several variables. For example,

$$q = q(x,y,z) = ax^2 + by^2 + cz^2 + 2hxy + 2fyz + 2gzx$$

In general, if there are n variables $x_1, x_2, \dots, x_n$ then a quadratic $q = q(x_1, x_2, \dots, x_n)$ is given by

$$q = a_{11}x_1^2 + a_{22}x_2^2 + \dots + a_{nn}x_n^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + \cdots 2a_{n-1n}x_{n-1}x_n$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

where $a_{ij} = a_{ji}$.

Then q can be written as $q = X^t A X$ where

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{pmatrix}, \qquad A = (a_{ij})_{n\times n}$$

Thus, $q = ax^2 + by^2 + cz^2 + 2hxy + 2fyz + 2gzx$ can be written as

$$q = (x\ y\ z)\begin{pmatrix} a & h & g \\ h & b & f \\ g & f & c \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = (x \quad y \quad z)\begin{pmatrix} ax + hy + gz \\ hx + by + fz \\ gx + fy + cz \end{pmatrix}$$

$$= ax^2 + by^2 + cz^2 + 2hxy + 2fyz + 2gzx$$

Remember, here we took, $a_{11} = a, x_1 = x, a_{22} = b, x_2 = y,\ a_{33} = c, x_3 = z, a_{12} = a_{21} = h,\ a_{23} = a_{32} = f$ and $a_{13} = a_{31} = g$.

Classification of quadratic forms

Suppose we have two quadratic forms. (1) $2x_1^2 + 3x_2^2$and(2)$x_1^2 + x_2^2 - 2x_1x_2$. Look, both have same range set, that is, the set of all non-negative real numbers. But note the difference, (1) gives the value 0 only when $x_1 = x_2 = 0$ whereas (2) gives zero if and only if $x_1 = x_2$, that is, x_1, x_2 may not be 0. Thus, we see that if

(1) $q = 2x_1^2 + 3x_2^2 = X^tAX$ where $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ then $q > 0$ for all $X \neq 0$, $X \in \mathbb{R}^2$

Here 0 means the null vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

(2) $q = x_1^2 + x_2^2 - 2x_1x_2 = X^tAX$ where $A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ then $q \geq 0$ for all $X \in \mathbb{R}^2$ and $q = 0$ for some $X \neq 0$. For example, if we take $X = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ then $X \neq 0$ but $q = 0$.

(3) $q = -2x_1^2 - 3x_2^2$ then $q < 0$ for all $X \neq 0, X \in \mathbb{R}^2$.

(4) $q = -x_1^2 - x_2^2 + 2x_1x_2$ then $q \leq 0$ for $X \in \mathbb{R}^2$ and $q = 0$ for some $X \neq 0$, e.g. $X = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

(5) $q = 2x_1^2 - 3x_2^2$ then $q > 0$ for some $X \neq 0$, e.g., $X = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $q < 0$ for some $X \neq 0$, e.g., $X = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Now, let us classify quadratic forms

2.7.3 Definition. A real quadratic form $q = X^tAX$ where $X \in \mathbb{R}^n$, $A \in M_{n\times n}(\mathbb{R})$ is called

(i) **positive definite** if $q > 0$ for all $X \neq 0$,

(ii) **positive semi definite** if $q \geq 0$ for all $X \in \mathbb{R}^n$ and $q = 0$ for some $X \neq 0$,

(iii) **negative definite** if $q < 0$ for all $X \neq 0$,

(iv) **negative semi definite** if $q \leq 0$ for all $X \in \mathbb{R}^n$ and $q = 0$ for some $X \neq 0$,

(v) **indefinite** if $q > 0$ for some $X \neq 0$ and $q < 0$ for some $X \neq 0$.

Note that every non-zero quadratic form belongs to exactly one of the categories: positive definite, positive semi definite, negative definite, negative semi definite and indefinite.Every quadratic form on $\mathbb{R}^n$ gives rise to a symmetric matrix and vice-versa. We say that the real symmetric matrix A is positive definite, positive semi definite, negative definite etc. if the associated quadratic form X^tAX is positive definite, positive semi definite, negative definite etc.

How to find the character of a given quadratic form?

There are several methods.

Method 1.

Sometimes it can be done by inspection together with some simple calculations. For ex. Let $q = x^2 + 2y^2 + z^2 + 2xy + 2yz$. Then

$$q = (x + y)^2 + (y + z)^2$$

So, $q \geq 0$ for all $(x, y, z) \in \mathbb{R}^3$. Now, $q = 0$ if $x + y = 0, y + z = 0$ i.e. if $x = -y = z$.

Thus, $q(1, -1, 1) = 0$ and $(1, -1, 1) \neq (0,0,0)$. Hence, q is positive semi definite.

But it is not easy to classify all quadratic forms by method 1. So, we need other methods.

Method 2

We know that a given real quadratic form q can be written as $q = X^tAX$ where A is a symmetric matrix. What happens if we use the transformation $X = PY$ where P is a non-singular matrix? Then we have,

$$q = (PY)^tA(PY) = Y^t(P^tAP)Y$$

Look, P^tAP is also a symmetric matrix as $(P^tAP)^t = P^tA^t(P^t)^t = P^tAP$ as $A^t = A$ and $(P^t)^t = P$.

Since $PY \neq 0$ iff $Y \neq 0$ it follows that X^tAX and $Y^t(P^tAP)Y$ have the same definiteness category. Thus a non-singular transformation of the variables changes a quadratic forminto another with the same definiteness category.

Hence, q becomes a quadratic form in Y. Now, we have learnt from 2.6 that for a given symmetric matrix A, there always exist a non-singular matrix Q, such that Q^tAQ is a diagonal matrix. Hence, with the help of a suitable matrix P, we can bring the given quadratic form $q(x_1, x_2, \ldots, x_n)$ to a form given by

$$y_1^2 + y_2^2 + \cdots + y_m^2 - y_{m-1}^2 - \cdots - y_r^2, \quad 0 \leq m \leq r \leq n$$

which is known as **normal form**or **canonical form**of q.

In normal form or in diagonal form of q if all diagonal elements are positive (resp. negative) then q is positive (res. negative) definite, if at least one of the diagonal elements is zero and others are positive (resp. negative) then q is positive (resp. negative) semi definite and if at least one of the diagonal elements is positive and at least one of the same is negative then q is indefinite.

Let $q(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + x_3^2 - 2x_1x_2 - 6x_1x_3 + 2x_2x_3$. Thus,

$$q = X^tAX \text{ where } X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & -1 & 3 \\ -1 & 2 & 1 \\ 3 & 1 & 1 \end{pmatrix}$$

Here A is a 3×3 real symmetric matrix. We start with a 3×6 matrix $(A|I)$ as follows:

$$(A|I) = \left(\begin{array}{ccc|ccc} 1 & -1 & 3 & 1 & 0 & 0 \\ -1 & 2 & 1 & 0 & 1 & 0 \\ 3 & 1 & 1 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{C_2+C_1}\left(\begin{array}{ccc|ccc}1&0&3&1&0&0\\-1&1&1&0&1&0\\3&4&1&0&0&1\end{array}\right)$$

$$\xrightarrow{R_2+R_1}\left(\begin{array}{ccc|ccc}1&0&3&1&0&0\\0&1&4&1&1&0\\3&4&1&0&0&1\end{array}\right)$$

$$\xrightarrow{C_3-3C_1}\left(\begin{array}{ccc|ccc}1&0&0&1&0&0\\0&1&4&1&1&0\\3&4&-8&0&0&1\end{array}\right)\xrightarrow{R_3-3R_1}\left(\begin{array}{ccc|ccc}1&0&0&1&0&0\\0&1&4&1&1&0\\0&4&-8&-3&0&1\end{array}\right)$$

$$\xrightarrow{C_3-4C_2}\left(\begin{array}{ccc|ccc}1&0&0&1&0&0\\0&1&0&1&1&0\\0&4&-24&-3&0&1\end{array}\right)\xrightarrow{R_3-4R_2}\left(\begin{array}{ccc|ccc}1&0&0&1&0&0\\0&1&0&1&1&0\\0&0&-24&-7&-4&1\end{array}\right)$$

$$\xrightarrow{\frac{1}{\sqrt{24}}C_3}\left(\begin{array}{ccc|ccc}1&0&0&1&0&0\\0&1&0&1&1&0\\0&0&-\sqrt{24}&-7&-4&1\end{array}\right)\xrightarrow{\frac{1}{\sqrt{24}}R_3}\left(\begin{array}{ccc|ccc}1&0&0&1&0&0\\0&1&0&1&1&0\\0&0&-1&-\frac{7}{\sqrt{24}}&-\frac{4}{\sqrt{24}}&\frac{1}{\sqrt{24}}\end{array}\right)$$

$$=(N|P^t)$$

Where

$$N=\begin{pmatrix}1&0&0\\0&1&0\\0&0&-1\end{pmatrix},\qquad P^t=\begin{pmatrix}1&0&0\\1&1&0\\-\frac{7}{\sqrt{24}}&-\frac{4}{\sqrt{24}}&\frac{1}{\sqrt{24}}\end{pmatrix}$$

and therefore, $P=\begin{pmatrix}1&1&-\frac{7}{\sqrt{24}}\\0&1&-\frac{4}{\sqrt{24}}\\0&0&\frac{1}{\sqrt{24}}\end{pmatrix}$

Hence, $P^tAP=N$ and the normal form of the given quadratic form is $y_1^2+y_2^2-y_3^2$ and therefore the form is indefinite.

We obtained the normal form by the substitution $X=PY$ i.e.

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}1&1&-\frac{7}{\sqrt{24}}\\0&1&-\frac{4}{\sqrt{24}}\\0&0&\frac{1}{\sqrt{24}}\end{pmatrix}\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}$$

That is, the required substitution is

$$x_1=y_1+y_2-\frac{7}{\sqrt{24}}y_3$$
$$x_2=y_2-\frac{4}{\sqrt{24}}y_3$$
$$x_3=\frac{1}{\sqrt{24}}y_3$$

Remark :For definiteness of the symmetric matrix A or the quadratic form $q=X^tAX$, it is not necessary to bring A into normal form, it is enough to bring it into a diagonal matrix and one can determine the definiteness of q by watching the diagonal elements only. If you are asked to find normal form then you have to bring A into a normal form, that is, to a diagonal form containing $1,-1$ and 0 only at the diagonal and in the normal form of q, negative signs come after all positive signs. For example, $y_1^2-y_2^2+y_3^2$ is not a normal form but $y_1^2+y_2^2-y_3^2$ is a normal form.

Method 3

We know, that eigen values (if you forget, please consult my book *Ring Theory and Linear Algebra*)of a symmetric matrix are all real. Thus, if $q=X^tAX$ where A is a symmetric matrix, we can find the eigenvalues $\lambda_1,\lambda_2,\dots,\lambda_n$. Then $q=X^tAX$ is positive (resp. negative) definite if $\lambda_i>0\ (resp.<0)$ for all $i=1,2,\dots,n$, is positive (resp. negative) semi definite if $\lambda_i\geq 0\ (resp.\leq 0)$ for all i and $\lambda_k=0$ for some $k\in\{1,2,\dots,n\}$. Now, q is indefinite if there exists some $\lambda_i>0$ and some $\lambda_j<0$.

Example. Let us consider the quadratic form $q=x^2+5y^2+2z^2-4xy-6yz+2zx$

Here, $q=X^tAX$ where $X=\begin{pmatrix}x\\y\\z\end{pmatrix}$, $A=\begin{pmatrix}1&-2&1\\-2&5&-3\\1&-3&2\end{pmatrix}$

The characteristic equation of A is $|A-\lambda I|=0$, i.e.

$$\begin{vmatrix}1-\lambda&-2&1\\-2&5-\lambda&-3\\1&-3&2-\lambda\end{vmatrix}=0\Rightarrow\ \lambda=0,4\pm\sqrt{13}$$

Since, at least one eigen value is 0 and others are positive, the given quadratic form is positive semi definite.

2.7.4 Theorem : A real symmetric matrix is positive definite if and only if all its eigenvalues are positive.

Proof. Let A be a real symmetric matrix of order n. Then all eigenvalues of A are real, say, $\lambda_1,\lambda_2,\dots,\lambda_n$.

Since A is a real symmetric matrix, there exists an orthogonal matrix P such that $P^{-1}AP$, equivalently, P^tAP, is a diagonal matrix.

Let $P^{-1}AP = diag\ (\lambda_1, \lambda_2, \dots, \lambda_n)$ where $\lambda_1, \lambda_2, \dots, \lambda_n$ are eigenvalues of A. Since, $P^tP = I$ and $\det P^{-1} \det P = 1$, we have,

$$\det(P^{-1}AP - \lambda I) = \det(P^{-1}AP - P^{-1}(\lambda I)P) = \det(P^{-1}(A - \lambda I)P)$$
$$= \det P^{-1} \det(A - \lambda I) \det P = \det(A - \lambda I)$$

Hence, the matrices A and $P^{-1}AP$ have same eigenvalues. In other words, $\lambda_1, \lambda_2, \dots, \lambda_n$ are eigenvalues of A.

Since, P is non-singular, $P^{-1}AP$ is congruent to A. If A is positive definite, then A is congruent to I_n. Therefore, $P^{-1}AP$ is congruent to I_n.

Hence, all $\lambda_i > 0$ for $i = 1, 2, \dots, n$.

Thus, if A is positive definite then all eigenvalues of A are positive.

Conversely, let all eigenvalues of A are positive, i.e, $\lambda_i > 0$ for $i = 1, 2, \dots, n$. Then $diag\ (\lambda_1, \lambda_2, \dots, \lambda_n)$ is positive definite.

But A is congruent to $P^{-1}AP = diag\ (\lambda_1, \lambda_2, \dots, \lambda_n)$.

Hence, A is positive definite ■

Note. Similarly, it can be proved that a real symmetric matrix is negative definite if and only if all its eigenvalues are negative.

Method 4

A real symmetric matrix $A = \begin{pmatrix} a & h & g \\ h & b & f \\ g & f & c \end{pmatrix}$

is positive definite if and only if

$$a > 0, \quad \begin{vmatrix} a & h \\ h & b \end{vmatrix} > 0 \ and \ \begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix} > 0$$

and is negative definite if and only if

$$a < 0, \quad \begin{vmatrix} a & h \\ h & b \end{vmatrix} > 0 \ and \ \begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix} < 0.$$

Example. For the matrix $A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -3 \\ 0 & -3 & 14 \end{pmatrix}$

we see that $2 > 0$, $\begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3 > 0$ and $\begin{vmatrix} 2 & -1 & 0 \\ -1 & 2 & -3 \\ 0 & -3 & 14 \end{vmatrix} = 24 > 0$

Hence, A is positive definite.

2.7.5 Definition. Let A be a real symmetric matrix, and let D be a diagonal matrix that is congruent to A. Let P denote the number of positive diagonal entries of D and N be the number of negative diagonal entries of D The number P is called index of A or of the quadratic form X^tAX and the number $P - N$ is called the signature of A or of the quadratic form X^tAX.

If R be the rank of the matrix A, then we have $P + N = R$ and hence signature is $P - N = P - (R - P) = 2P - R$. Some authors prefer to define signature as $2P - R$, that is, $2 \times Index - Rank$. Thus, if we denote signature of X^tAX, i.e. of A by S, we have,

$$S = P - N = 2P - R$$

Hence, we can say that, an n-ary quadratic form X^tAX, that is, A is an $n \times n$ symmetric matrix with rank R and signature S having P and N as defined above is

(i) ***Positive definite*** if $P = n$ (equivalently, $S = n$)
(ii) ***Positive semi definite*** if $N = 0, P = R < n$ (equivalently, $S = R < n$)
(iii) ***Negative definite*** if $P = 0, N = R = n$ (equivalently, $S = -n$)
(iv) ***Negative semi definite*** if $P = 0, N = R < n$ (equivalently $S = -R < n$)
(v) ***Indefinite*** if $P \geq 1$ and $N \geq 1$

2.8 SECOND DERIVATIVE TEST

It is interesting to show how the theory of quadratic forms be applied to multivariable calculus. But before that we need to understand the followings.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real valued function and we can express as $z = f(t)$ where $t = (t_1, t_2, \dots, t_n) \in \mathbb{R}^n$, i.e., $z = f(t_1, t_2, \dots, t_n)$.

Let all third-order partial derivatives of f exist and are continuous.

The function f is said to have a **local maximum** at a point $p \in \mathbb{R}^n$ if there exists a $\delta > 0$ such that $f(p) \geq f(x)$ whenever $\|x - p\| < \delta$. Similarly, f is said to have a **local minimum** at a point $p \in \mathbb{R}^n$ if there exists $\delta > 0$ such that $f(p) \leq f(x)$ whenever $\|x - p\| < \delta$. If f has either local minimum or a local maximum at p then we say that f has a **local extremum** at p.

A point $p \in \mathbb{R}^n$ is said to be a **critical point** of f if $\frac{\partial f(p)}{\partial t_i} = 0$ for $i = 1, 2, \dots, n$.

We know from calculus that if f has a local extremum at $p \in \mathbb{R}^n$, then p is a critical point of f. For, if f has a local extremum at $p = (p_1, p_2, \dots, p_n)$, then for any $i = 1,2,\dots,n$ the function ϕ_i defined by $\phi_i(t) = f(p_1, p_2, \dots, p_{i-1}, t, p_{i+1}, \dots, p_n)$ has a local extremum at $t = p_i$. Thus, we have,

$$\frac{\partial f(p)}{\partial t_i} = \frac{d\phi_i(p_i)}{dt} = 0.$$

Therefore, p is a critical point of f. But critical points are not necessarily local extrema.

The second order partial derivatives of f at a critical point p can often be used to test for a local extremum at p. These partials determine a matrix $A(p) = (a_{ij})_{n\times n}$ where

$$a_{ij} = \frac{\partial^2 f(p)}{(\partial t_i)(\partial t_j)}$$

This matrix is called the Hessian matrix of f at p.

Since the third order partial derivatives of f are continuous, then

$$\frac{\partial^2 f(p)}{(\partial t_i)(\partial t_j)} = \frac{\partial^2 f(p)}{(\partial t_j)(\partial t_i)}, \qquad i.e.\ a_{ij} = a_{ji}$$

Hence, the Hessian matrix $A(p)$ of f at p is a symmetric matrix. Thus, all eigen values of $A(p)$ are real.

Example. Let us consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x,y) = x^2 + 3xy + y^2$ and $p = (1,2)$. We wish to have Hessian matrix $A(p)$ of f at p.

Here, $f_x(x,y) = 2x + 3y,\ f_y(x,y) = 3x + 2y,\ f_{xy}(x,y) = 3 = f_{yx}(x,y)$.

So, $f_x(1,2) = 8,\ f_y(1,2) = 7,\ f_{xy}(x,y) = 3 = f_{yx}(x,y)$.

Thus, $A(p) = \begin{pmatrix} 8 & 3 \\ 3 & 7 \end{pmatrix}$.

2.8.1 Theorem (The Second Derivative Test) Let $f(t_1, t_2, \dots, t_n)$ be a real valued function in n real variables for which all third order partial derivatives exists and are continuous. Let $p = (p_1, p_2, \dots, p_n)$ be a critical point of f, and let $A(p)$ be the Hessian of f at p.

(a) If all eigenvalues of $A(p)$ are positive (i.e. $A(p)$ is positive definite) then f has a local minimum at p.

(b) If all eigenvalues of $A(p)$ are negative (i.e. $A(p)$ is negative definite) then f has a local maximum at p.

(c) If $A(p)$ has at least one positive and at least one negative eigenvalue (i.e. $A(p)$ is indefinite) then f has no local extremum at p. (In this case, p is called a saddle point of)

(d) If $rank(A(p)) < n$ and $A(p)$ does not have both positive or negative eigenvalues (i.e. if $A(p)$ is positive semi definite or negative semi definite) then the second derivative test is inconclusive.

Proof. If $p \neq 0$, that is $p = (p_1, p_2, \dots, p_n)$ where all p_i's are not zero, let us define $g : \mathbb{R}^n \to \mathbb{R}$ by

$$g(t_1, t_2, \dots, t_n) = f(t_1 + p_1, t_2 + p_2, \dots, t_n + p_n) - f(p_1, p_2, \dots, p_n)$$

It is evident that,

1. The function f has a local minimum (maximum) at p if and only if g has a local minimum (maximum) at $0 = (0,0,\dots,0)$.
2. The partial derivatives of g at 0 are equal to the corresponding partial derivatives of f at p.
3. $g(0) = 0$
4. 0 is a critical point of g, i.e. $\frac{\partial g(0)}{\partial t_i} = 0$ for all $i = 1,2,\dots,n$
5. The Hessian matrix $A(p) = (a_{ij})_{n\times n}$ where

$$a_{ij} = \frac{\partial^2 f(p)}{(\partial t_i)(\partial t_j)} = \frac{\partial^2 g(0)}{(\partial t_i)(\partial t_j)}.$$

Let us apply Taylor's theorem to g around 0 and we get

$$g(t_1, t_2, \dots, t_n) = g(0) + \sum_{i=1}^{n} \frac{\partial g(0)}{\partial t_i} t_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 g(0)}{(\partial t_i)(\partial t_j)} t_i t_j + S(t_1, t_2, \dots, t_n) \ \dots\dots (1)$$

where S is a real valued function on $\mathbb{R}^n$ (i.e. $S : \mathbb{R}^n \to \mathbb{R}$) such that

$$\lim_{x\to 0} \frac{S(x)}{\|x\|^2} = \lim_{(t_1,t_2,\dots,t_n)\to 0} \frac{S(t_1, t_2, \dots, t_n)}{t_1^2 + t_2^2 + \dots + t_n^2} = 0 \ \dots\dots (2)$$

Since, $g(0) = 0$ and $\frac{\partial g(0)}{\partial t_i} = 0$ for all $i = 1,2,\dots,n$, equation (1) becomes

$$g(t_1, t_2, \dots, t_n) = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 g(0)}{(\partial t_i)(\partial t_j)} t_i t_j + S(t_1, t_2, \dots, t_n) \ \dots\dots (3)$$

Let $K : \mathbb{R}^n \to \mathbb{R}$ be the quadratic form defined by

$$K\begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix} = \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 g(0)}{(\partial t_i)(\partial t_j)} t_1 t_j \;\ldots\ldots (4)$$

Let H be the symmetric bilinear form corresponding to K, and β be the standard ordered basis for $\mathbb{R}^n$. Then H has the matrix representation $\psi_\beta(H) = (H_{ij})_{n\times n}$ where

$$H_{ij} = \frac{1}{2}\frac{\partial^2 g(0)}{(\partial t_i)(\partial t_j)} = \frac{1}{2}a_{ij}$$

Hence, $\psi_\beta(H) = \frac{1}{2}A(p)$.

Since, $A(p)$ is symmetric there exists an orthogonal matrix Q such that

$$Q^t A(p) Q = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

is a diagonal matrix whose diagonal entries are the eigenvalues of $A(p)$. Let $\gamma = \{v_1, v_2, \ldots, v_n\}$ be the orthogonal basis for $\mathbb{R}^n$ whose ith vector is the ith column of Q. Then Q is the change of coordinate matrix changing γ −coordinates into β − coordinates and hence,

$$\psi_\gamma(H) = Q^t\psi_\beta(H)Q = \frac{1}{2}Q^t A(p) Q = \begin{pmatrix} \frac{\lambda_1}{2} & 0 & \ldots & 0 \\ 0 & \frac{\lambda_2}{2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \cdots & \frac{\lambda_n}{2} \end{pmatrix}$$

Suppose that $A(p)$ is not the zero matrix. Then $A(p)$ has non-zero eigenvalues. Let us choose $\epsilon > 0$ such that $\epsilon < \frac{|\lambda_i|}{2}$ for all $\lambda_i \neq 0$. Now, by (2),

$$\lim_{x\to 0}\frac{S(x)}{\|x\|^2} = 0$$

So, there exists $\delta > 0$ such that for any $x \in \mathbb{R}^n$ satisfying $0 < \|x\| < \delta$, we have, $\frac{|S(x)|}{\|x\|^2} < \epsilon$, that is, $|S(x)| < \epsilon.\|x\|^2$. Consider any $x \in \mathbb{R}^n$ such that $0 < \|x\| < \delta$. Then by (3) and (4)

$$|g(x) - K(x)| = |S(x)| < \epsilon\|x\|^2,$$

$$\text{i.e.} \quad -\epsilon\|x\|^2 < g(x) - K(x) < \epsilon\|x\|^2$$

$$\text{i.e.} \quad K(x) - \epsilon\|x\|^2 < g(x) < K(x) + \epsilon\|x\|^2 \;\ldots (5)$$

Let $x = \sum_{i=1}^{n} s_i v_i$ where $s_1, s_2, \ldots, s_n$ are scalars. Now,

$\|x\|^2 = <x, x> = <s_1v_1 + s_2v_2 + \cdots + s_nv_n, s_1v_1 + s_2v_2 + \cdots + s_nv_n>$

$= s_1^2 + s_2^2 + \cdots + s_n^2 [\,as <v_i, v_j> = 0 \; for \; i \neq j, <v_i, v_i> = 1\,]$

$$= \sum_{i=1}^{n} s_i^2$$

and $K(x) = \frac{1}{2}\sum_{i=1}^{n}\lambda_i s_i^2$.

Thus, by (5), we have,

$$\sum_{i=1}^{n}\frac{1}{2}\lambda_i s_i^2 - \epsilon\sum_{i=1}^{n} s_i^2 < g(x) < \sum_{i=1}^{n}\frac{1}{2}\lambda_i s_i^2 + \epsilon\sum_{i=1}^{n} s_i^2$$

$$\Rightarrow \sum_{i=1}^{n}\left(\frac{1}{2}\lambda_i - \epsilon\right)s_i^2 < g(x) < \sum_{i=1}^{n}\left(\frac{1}{2}\lambda_i + \epsilon\right)s_i^2 \;\ldots\ldots (6)$$

If all eigenvalues of $A(p)$ are positive then $\frac{1}{2}\lambda_i - \epsilon > 0$ for all i, and hence by left inequality in (6)

$$g(0) = 0 \le \sum_{i=1}^{n}\left(\frac{1}{2}\lambda_i - \epsilon\right)s_i^2 < g(x).$$

Thus, $g(0) \le g(x)$for $\|x\| < \delta$, and so g has a local minimum at 0 and hence f has a local minimum at p. by a similar argument using the right inequality in (6), we have that if all of the eigenvalues of $A(p)$ are negative, then g has a local maximum at 0, that is, f has a local maximum at p. hence(a) and (b) of the theorem are proved.

Now, let $A(p)$ has both a positive and a negative eigenvalue, say, $\lambda_i > 0$ and $\lambda_j < 0$ for some $i$and j. Then $\frac{1}{2}\lambda_i - \epsilon > 0and\frac{1}{2}\lambda_j + \epsilon < 0$. Let s be any real number satisfying $0 < |s| < \delta$. Substituting $x = sv_i$ and $x = sv_j$ into the left inequality and the right inequality of (6), respectively, we get

$$g(0) = 0 < \left(\frac{1}{2}\lambda_i - \epsilon\right)s^2 < g(sv_i)$$

and $g(sv_j) < \left(\frac{1}{2}\lambda_j + \epsilon\right)s^2 < 0 = g(0).$

Thus, g attains both positive and negative values arbitrarily close to 0; so g has neither a local maximum nor a local minimum at 0 which shows that f has neither a local maximum nor a local minimum at p. Hence, (c) is proved.

For the last part (d) of the theorem, let us consider the functions

$$F(t_1,t_2)=t_1^2-t_2^4, \quad G(t_1,t_2)=t_1^2+t_2^4$$

at $p=0$. In both cases, the function has a critical point at p, and

$$A(p)=\begin{pmatrix}2&0\\0&0\end{pmatrix}.$$

However, F does not have a local extremum at 0 whereas G has a local minimum at 0 ■

Example. Let $f:\mathbb{R}^3\to\mathbb{R}$ be a function given by

$$f(x,y,z)=x^2+y^2+7z^2-xy-3yz$$

First, we set the partial derivatives of first order equal to 0

$$f_x(x,y,z)=2x-y=0, \quad f_y(x,y,z)=2y-x-3z=0,$$
$$f_z(x,y,z)=14z-3y=0$$

Since, $\begin{vmatrix}2&-1&0\\-1&2&-3\\0&-3&14\end{vmatrix}\neq 0$, we see that $(0,0,0)$ is the only solution of

$$f_x(x,y,z)=f_y(x,y,z)=f_z(x,y,z)=0$$

Hence, $(0,0,0)$ is the only critical point of f.

Now,

$$f_{xx}(x,y,z)=2, \quad f_{xx}(0,0,0)=2$$
$$f_{yy}(x,y,z)=2, \quad f_{yy}(0,0,0)=2$$
$$f_{zz}(x,y,z)=14, \quad f_{zz}(0,0,0)=14$$
$$f_{xy}(x,y,z)=-1, \quad f_{xy}(0,0,0)=-1$$
$$f_{yz}(x,y,z)=-3, \quad f_{yz}(0,0,0)=-3$$
$$f_{zx}(x,y,z)=0, \quad f_{zx}(0,0,0)=-3$$

Writing $p=(1,1,1)$, we have, the Hessian matrix $A(p)$ of f at p as

$$A(p)=\begin{pmatrix}f_{xx}(p)&f_{xy}(p)&f_{xz}(p)\\f_{yx}(p)&f_{yy}(p)&f_{yz}(p)\\f_{zx}(p)&f_{zy}(p)&f_{zz}(p)\end{pmatrix}=\begin{pmatrix}2&-1&0\\-1&2&-3\\0&-3&14\end{pmatrix}$$

Let us apply congruence operation on $A(p)$ to get a diagonal matrix

$$\xrightarrow{R_2+\frac{1}{2}R_1}\begin{pmatrix}2&-1&0\\0&\frac{3}{2}&-3\\0&-3&14\end{pmatrix}\xrightarrow{C_2+\frac{1}{2}C_1}\begin{pmatrix}2&0&0\\0&\frac{3}{2}&-3\\0&-3&14\end{pmatrix}$$

$$\xrightarrow{R_3+2R_2}\begin{pmatrix}2&0&0\\0&\frac{3}{2}&-3\\0&0&8\end{pmatrix}\xrightarrow{C_3+2C_2}\begin{pmatrix}2&0&0\\0&\frac{3}{2}&0\\0&0&8\end{pmatrix}$$

Since, all diagonal entries are positive, the matrix $A(p)$ is positive definite and hence by the second derivative test we can say that f has a local minimum at $(0,0,0)$.

Alternatively,

$$A(p)=\begin{pmatrix}2&-1&0\\-1&2&-3\\0&-3&14\end{pmatrix}$$

we see that $2>0$, $\begin{vmatrix}2&-1\\-1&2\end{vmatrix}=3>0$ and $\begin{vmatrix}2&-1&0\\-1&2&-3\\0&-3&14\end{vmatrix}=24>0$

and hence the matrix $A(p)$ is positive definite and therefore, by the second derivative test we can say that f has a local minimum at $(0,0,0)$.

2.9 SYLVESTER'S LAW OF INERTIA

We know that, by a suitable transformation $X=PY$, where P is a non-singular matrix, the real quadratic form $q=X^tAX$ can be brought into a diagonal form like

$$\lambda_1y_1^2+\lambda_2y_2^2+\cdots+\lambda_ry_m^2-\lambda_{r+1}y_{r+1}^2-\cdots-\lambda_ry_r^2+0.y_{r+1}^2+\cdots+0.y_n^2$$

or in normal form like

$$y_1^2+y_2^2+\cdots+y_m^2-y_{m+1}^2-\cdots-y_r^2+0.y_{r+1}^2+\cdots+0.y_n^2$$

Remember, normal form is also a diagonal form whose diagonal entries can take value only $1.-1$and 0, nothing else. Now, the number r, the number of non-zero entries in the diagonal matrix congruent to A is called the *rank* and m, the number of positive entries in the diagonal matrix congruent to A is called the *index* of the real quadratic form and we also know that $s=2m-r$ is the *signature* of q. Now, the question is whether these numbers m,r,s are uniquely determined for a given quadratic form q. In other words, if we apply another transformation $X=UZ$, where U is a non-singular matrix, and we get a diagonal form like

$$\mu_1 z_1^2 + \mu_2 z_2^2 + \cdots + \mu_k z_k^2 - \mu_{k+1} z_{k+1}^2 - \cdots - \mu_r z_r^2 + 0.z_{r+1}^2 + \cdots 0.z_n^2$$

then can we say that $m = k$? The answer lies in the following theorem.

2.9.1 Theorem. The number of positive elements in the normal (diagonal) form of a real quadratic form is invariant.

Proof. Without any loss of generality, we consider only normal forms.

Let $q = X^t AX$ be a real quadratic form where $X = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}$

Let $X = BY$, B being a non-singular matrix, be the substitution which transforms q to the normal form

$$y_1^2 + y_2^2 + \cdots + y_m^2 - y_{m+1}^2 - \cdots - y_r^2 + 0.y_{r+1}^2 + \cdots + 0.y_n^2.$$

Let $X = CZ$, C being a non-singular matrix, be another substitution which transforms q to the normal form

$$z_1^2 + z_2^2 + \cdots + z_k^2 - z_{k+1}^2 - \cdots - z_r^2 + 0.z_{r+1}^2 + \cdots + 0.z_n^2.$$

We claim that $m = k$. If not, let $m < k$.

Now, $Y = B^{-1}X$ and $Z = C^{-1}X$. Let $B^{-1} = (b_{ij})_{n\times n}$, $C^{-1} = (c_{ij})_{n\times n}$.

Thus we have,

$$y_j = b_{j1}x_1 + b_{j2}x_2 + \cdots + b_{jn}x_n, \quad for \ j = 1,2,..,n$$

and

$$z_j = c_{j1}x_1 + c_{j2}x_2 + \cdots + c_{jn}x_n \quad for \ j = 1,2,\ldots,n \ldots\ldots (1)$$

Let us consider $m + n - k$ equations in n unknowns

$$b_{11}x_1 + b_{12}x_2 + \cdots + b_{1n}x_n = 0$$

$$\ldots\ldots\ldots$$

$$b_{m1}x_1 + b_{m2}x_2 + \cdots + b_{mn}x_n = 0$$

$$c_{k+11}x_1 + c_{(k+1)2}x_2 + \cdots + c_{k+1\,n}x_n = 0$$

$$\ldots\ldots\ldots$$

$$c_{n1}x_1 + c_{n2}x_2 + \cdots + c_{nn}x_n = 0$$

Since, $m < k \Rightarrow m - k < 0 \Rightarrow m + n - k < n$, the above system of equations has non-zero solutions. Let $X' = (x_1', x_2', \ldots, x_n')$ be such a non-zero solution.

When $X = X'$ let $Y = Y' = (y_1', y_2', \ldots, y_n')$ and $Z = Z' = (z_1', z_2', \ldots, z_n')$.

Then first m equations together with (1) show that $y_1' = y_2' = \cdots y_m' = 0$ and last $n - k$ equations together with (1) show that $z_{k+1}' = z_{k+2}' = \cdots = z_n' = 0$.

Thus, we have,

$$q(x_1', x_2', \ldots, x_n') = -y_{m+1}'^2 - y_{m+2}'^2 - \cdots - y_r'^2 = z_1'^2 + z_2'^2 + \cdots + z_k'^2$$

It happens iff $y_{m+1}' = y_{m+2}' = \cdots = y_r' = 0$ and $z_1' = z_2' = \cdots = z_k' = 0$.

Therefore, we get $Y' = 0 = (0,0,\ldots,0)$, $Z' = 0 = (0,0,\ldots,0)$.

Since, B is non-singular, we have, $X' = BY' = 0 = (0,0,\ldots,0)$ which contradicts that X' is non-zero.

This contradiction shows that $m \not< k$. Similarly, it can be shown that $k \not< m$.

Hence, $m = k$.

Therefore, m is invariant for a given quadratic form q ■

Corollary: Since the rank r is invariant under congruence, the signature $s = 2m - r$ is also invariant.

Solved Problems:

1. **Let $V = C[0,1]$ be the space of continuous real valued functions on the closed interval $[0,1]$. For $f, g \in V$, define**

$$H(f,g) = \int_0^1 f(t)g(t)dt$$

Is H a bilinear form on V ?

Solution. Let $f_1, f_2, g \in V$ and $c \in F$. Clearly, $cf_1 + f_2 \in V$. Now,

$$H(cf_1 + f_2, g) = \int_0^1 (cf_1 + f_2)(t)g(t)\,dt$$

$$= \int_0^1 (cf_1(t) + f_2(t))g(t)\,dt$$

$$= c\int_0^1 f_1(t)g(t)\,dt + \int_0^1 f_2(t)(t)g(t)\,dt$$

$$= cH(f_1, g) + H(f_2, g)$$

Similarly, for $f, g_1, g_2 \in V$ and $d \in F$, we have,

$$H(f, dg_1 + g_2) = \int_0^1 f(t)(dg_1 + g_2)(t)\,dt$$

$$= \int_0^1 f(t)(dg_1(t) + g_2(t))\,dt$$

$$= d\int_0^1 f(t)g_1(t)\,dt + \int_0^1 f(t)g_2(t)\,dt$$
$$= d\,H(f,g_1) + H(f,g_2)$$

Hence, H is a bilinear form on V.

2. Define $H : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ by $H(t_1,t_2) = t_1 + 2t_2$. Is H a bilinear form on $\mathbb{R}$?

Solution. We take $(1,1) \in \mathbb{R}\times\mathbb{R}$ and $2 \in \mathbb{R}$. Then

$$H(2.1,1) = H(2,1) = 2 + 2.1 = 4$$

but $2H(1,1) = 2(1+2.1) = 6$.

Thus, $H(2.1,1) \neq 2H(1,1)$. Hence. H is not a bilinear form on $\mathbb{R}$.

3. Prove that the sum of two bilinear forms is a bilinear form.

Solution. Let H_1 and H_2 be two bilinear forms defined on a vector space V over the field F. Let us define $H : V\times V \to F$ by $H(x,y) = H_1(x,y) + H_2(x,y)$. We shall show that H is bilinear on V.

Let $x_1, x_2, y \in V$ and $c \in F$. Then

$$H(cx_1 + x_2, y) = H_1(cx_1+x_2,y) + H_2(cx_1+x_2,y)$$
$$= cH_1(x_1,y) + H_1(x_2,y) + cH_2(x_1,y) + H_2(x_2,y)$$

$[\,as\ H_1, H_2\ are\ bilinear]$

$$= c[H_1(x_1,y) + H_2(x_1,y)] + H_1(x_2,y) + H_2(x_2,y)$$
$$= c\,H(x_1,y) + H(x_2,y)$$

Similarly, for $x, y_1, y_2 \in V$ and $d \in F$, it can be shown that

$$H(x, dy_1 + y_2) = dH(x,y_1) + H(x,y_2)$$

Hence, $H = H_1 + H_2$ is a bilinear form on V.

4. Let V and W be vector spaces over the same field F and let $T : V \to W$ be a linear transformation. For any $H \in \mathcal{B}(W)$[(W) being the set of all bilinear forms on W], define $\hat{T}(H) : V\times V \to F$ by $\hat{T}(H)(x,y) = H(T(x),T(y))$ for all $x,y \in V$. Prove that if $H \in \mathcal{B}(W)$ then $\hat{T}(H)$ $\mathcal{B}(V)$.

Solution. Let $x_1, x_2, y \in V$ and $c \in F$. Then

$$\hat{T}(H)(cx_1 + x_2, y)$$
$$= H(T(cx_1+x_2), T(y))$$
$$= H\,(\,cT(x_1) + T(x_2), T(y))[\,by\ linearity\ of\ T]$$

$$= cH(T(x_1),T(y)) + H(T(x_2),T(y))[\,bilinearity\ of\ H]$$
$$= c\hat{T}(H)(x_1,y) + \hat{T}(H)(x_2,y)$$

Again, for $x, y_1, y_2 \in V$ and $d \in F$, we have,

$$\hat{T}(H)(x, dy_1 + y_2)$$
$$= H(T(x), T(dy_1 + y_2))$$
$$= H(T(x), dT(y_1) + T(y_2))$$
$$= d\,H(T(x),T(y_1)) + H(T(x),T(y_2))$$
$$= d\hat{T}(H)(x,y_1) + \hat{T}(H)(x,y_2)$$

Hence, $\hat{T}(H)$ is bilinear on V, in other words, $\hat{T}(H) \in \mathcal{B}(V)$.

5. Let V be an n –dimensional vector space over a field F and $\mathcal{B}(V)$ be the set of all bilinear forms on V. For $H \in \mathcal{B}(V)$, let $\psi_B(H)$ be the matrix representation of H with respect to the ordered basis B for V. Then prove that $\psi_B(H)$ is linear for any ordered basis B.

Solution. Let $B = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis for V.

Let $\psi_B(H) = M = (m_{ij})_{n\times n}$

Then we have, $(\psi_B(H))_{ij} = m_{ij} = H(v_i, v_j)$.

For $H_1, H_2 \in \mathcal{B}(V)$ and for $c \in F$, we have,

$$(\psi_B(cH_1 + H_2))_{ij} = (cH_1 + H_2)(v_i,v_j)$$
$$= cH_1(v_i,v_j) + H_2(v_i,v_j)$$
$$= c\,(\psi_B(H_1))_{ij} + (\psi_B(H_2))_{ij}$$

Hence, $\psi_B(H)$ is linear.

6. Prove that (i) any square diagonal matrix is symmetric and (ii) any matrix congruent to a diagonal matrix is symmetric.

Solution. (i)Let $A = (a_{ij})_{n\times n}$ be a diagonal matrix. Then

$$a_{ij} = a_{ji} = 0\ \ for\ i \neq j.$$

Hence, A is a symmetric matrix.

(ii) Let A be a matrix congruent to a diagonal matrix B. Then there exists a invertible matrix Q such that $B = Q^t AQ$, that is $(Q^t)^{-1}BQ^{-1} = A$. Now,

$$A^t = [(Q^t)^{-1}BQ^{-1}]^t = (Q^{-1})^t B^t [(Q^t)^{-1}]^t$$
$$= (Q^t)^{-1}BQ^{-1}[\text{ as } B^t = B \text{ by } (i)] = A.$$

Hence, A is symmetric.

7. Let V be a vector space over a field F not of characteristic two, and let H be a symmetric bilinear form on V. Prove that if $K(x) = H(x,x)$ is the quadratic form associated with H, then show that for all $x, y \in V$,

$$H(x,y) = \frac{1}{2}[K(x+y) - K(x) - K(y)].$$

Solution. Using bilinear property of H, we have,

$$K(x+y) = H(x+y, x+y)$$
$$= H(x,x) + H(y,x) + H(x,y) + H(y,y)$$
$$= H(x,x) + 2H(x,y) + H(y,y)$$
$$[\text{as } H(x,y) = H(y,x), \quad H \text{ being symmetric}]$$

Since F is not of characteristic two, we have, $2H(x,y) \neq 0$ for arbitrary$(x,y) \in V \times V$. Hence,

$$H(x,y) = \frac{1}{2}[K(x+y) - K(x) - K(y)].$$

8. Let $K : \mathbb{R}^2 \to \mathbb{R}$ defined by $K\begin{pmatrix}a\\b\end{pmatrix} = -2a^2 + 4ab + b^2$ be a real quadratic form. Find a symmetric bilinear form H such that $K(x) = H(x,x)$ for all$x = \begin{pmatrix}a\\b\end{pmatrix} \in \mathbb{R}^2$.

Solution. Let $x = \begin{pmatrix}a\\b\end{pmatrix}$, $y = \begin{pmatrix}c\\d\end{pmatrix} \in \mathbb{R}^2$. Then using the formula

$$H(x,y) = \frac{1}{2}[K(x+y) - K(x) - K(y)]$$

We have,$H(x,y) = \frac{1}{2}\left[K\begin{pmatrix}a+c\\b+d\end{pmatrix} - K\begin{pmatrix}a\\b\end{pmatrix} - K\begin{pmatrix}c\\d\end{pmatrix}\right]$

$$= \frac{1}{2}[-2(a+c)^2 + 4(a+c)(b+d) + (b+d)^2 - (-2a^2 + 4ab + b^2)$$
$$- (-2c^2 + 4cd + d^2)]$$
$$= -2ac + 2ad + 2bc + bd.$$

Hence, the required symmetric bilinear form H is given by

$$H\left(\begin{pmatrix}a\\b\end{pmatrix}, \begin{pmatrix}c\\d\end{pmatrix}\right) = -2ac + 2ad + 2bc + bd.$$

9. Prove the following variation of the second derivative test for the case $n = 2$. Define

$$D = \left[\frac{\partial^2 f(p)}{\partial t_1^2}\right]\left[\frac{\partial^2 f(p)}{\partial t_2^2}\right] - \left[\frac{\partial^2 f(p)}{\partial t_1 \partial t_2}\right]^2$$

(a) If $D > 0$and $\frac{\partial^2 f(p)}{\partial t_1^2} > 0$, then f has a local minimum at p

(b) If $D > 0$and $\frac{\partial^2 f(p)}{\partial t_1^2} < 0$, then f has a local maximum at p

(c) If $D < 0$, then f has no local extremum at p

(d) If $D = 0$, then the test is inconclusive.

Solution.

(a) For $n = 2$, the Hessian matrix $A(p)$of f at p is given by

$$A(p) = \begin{pmatrix} \frac{\partial^2 f(p)}{\partial t_1^2} & \frac{\partial^2 f(p)}{\partial t_1 \partial t_2} \\ \frac{\partial^2 f(p)}{\partial t_1 \partial t_2} & \frac{\partial^2 f(p)}{\partial t_2^2} \end{pmatrix}$$

By the problem, we have, $D = \det A(p) = \frac{\partial^2 f(p)}{\partial t_1^2}\frac{\partial^2 f(p)}{\partial t_2^2} - \left[\frac{\partial^2 f(p)}{\partial t_1 \partial t_2}\right]^2$.

Let eigenvalues of $A(p)$ be λ_1 and λ_2.

Since, $D = \det A(p) > 0$and $\frac{\partial^2 f(p)}{\partial t_1^2} > 0$,

we have, $\frac{\partial^2 f(p)}{\partial t_2^2} > 0$, otherwise, D could be ≤ 0. Hence,

$$\lambda_1 + \lambda_2 = trace\ A(p) = \left[\frac{\partial^2 f(p)}{\partial t_1^2} + \frac{\partial^2 f(p)}{\partial t_2^2}\right] > 0$$

Again, $\lambda_1\lambda_2 = \det A(p) = D > 0$

This is possible, only when both λ_1 and λ_2 are positive. Hence, by the second derivative test we can say that f has a local minimum at p.

(b) If $D > 0$ and $\frac{\partial^2 f(p)}{\partial t_1^2} < 0$ then $\frac{\partial^2 f(p)}{\partial t_2^2}$ must be less than zero, otherwise, D would be negative. Hence,

$$\lambda_1 + \lambda_2 = trace\ A(p) < 0$$

and$\lambda_1\lambda_2 = \det A(p) = D > 0$

This is possible, only when both λ_1 and λ_2 are negative. Hence, by the second derivative test, we can say that f has a local maximum at p.

(c) If $D < 0$ then we have, $\lambda_1 \lambda_2 = \det A(p) = D < 0$.

In this case, one of $\lambda_i' s, i = 1,2$ must be positive and other one is negative. Hence, by the second derivative test, we see that, f has no local extremum at p.

(d) If $D = 0$, then by the relation, $\lambda_1 \lambda_2 = D = 0$

i.e.$rank\ A(p) < 2$. Hence, by the second derivative test, the test is inconclusive.

Exercise

1. Prove that the product of a scalar and a bilinear form is a bilinear form. In other words, if H is a bilinear form on a vector space V over a field $F$and $c \in F$, then prove that cH is a bilinear form on V.
2. Prove that the relation of congruence is an equivalence relation.
3. Let V be a finite dimensional vector space and L_1, L_2 linear functionals on V. Show that the equation
$$f(u, v) = L_1(u)L_2(v) - L_1(v)L_2(u)$$
defines a skew symmetric bilinear form on V. Show that $f = 0$ if and only if L_1, L_2 are linearly dependent.
4. Let f be any skew symmetric bilinear form on $\mathbb{R}^3$. Prove that there are linear functionals L_1, L_2 such that
$$f(u, v) = L_1(u)L_2(v) - L_1(v)L_2(u)$$
for all $u, v \in \mathbb{R}^3$.
5. Let $K : \mathbb{R}^2 \to \mathbb{R}$ defined by $K\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = 7t_1^2 - 8t_1t_2 + t_2^2$, be a real quadratic form. Find a symmetric bilinear form H such that $K(x) = H(x, x)$ for all $x = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \in \mathbb{R}^2$.
6. Let $K : \mathbb{R}^3 \to \mathbb{R}$ defined by $K\begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = 3t_1^2 + 3t_2^2 + 3t_3^2 - 2t_1t_3$ be a real quadratic form. Find a symmetric bilinear form H such that $K(x) = H(x, x)$ for all $x = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \in \mathbb{R}^3$.

1.10 DUAL SPACES

Let V be a vector space over a field F. Any linear mapping from $f : V \to F$ is called a linear functional on V. Thus, we have, the following definition.

1.10.1 Definition. Let V be a vector space over a field F. A mapping f from V to F is called a linear functional on V if it satisfies following property
$$f(cv_1 + v_2) = cf(v_1) + f(v_2)$$
for all vectors v_1 and v_2 in V and all scalars c in F.

1.10.2 Examples.

1. Let F be a field and let $a_1, a_2, \ldots, a_n$ be scalars in F. Define a function $f : F^n \to F$ by
$$f(v) = a_1v_1 + a_2v_2 + \cdots + a_nv_n$$
where $v = (v_1, v_2, \ldots, v_n)$. Let us check whether f is a linear functional on V.

Let $v = (v_1, v_2, \ldots, v_n)$ and $w = (w_1, w_2, \ldots, w_n)$ be two elements of F^n and $c \in F$. Then
$$\begin{aligned} f(cv + w) &= f(cv_1 + w_1, cv_2 + w_2, \ldots, cv_n + w_n) \\ &= a_1(cv_1 + w_1) + a_2(cv_2 + w_2) + \cdots + a_n(cv_n + w_n) \\ &= c(a_1v_1 + a_2v_2 + \cdots + a_nv_n) + (w_1 + w_2 + \cdots + w_n) \\ &= cf(v) + f(w) \end{aligned}$$

So, f is linear and hence is a linear functional on F^n as co-domain set of f is the scalar field F.

If $\{e_1, e_2, \ldots, e_n\}$ be the standard basis of F^n, that is, e_i is a vector in F^n whose $i\ th$ co-ordinate is 1 and all other coordinates are 0, then we have,
$$\begin{aligned} v &= (v_1, v_2, \ldots, v_n) \\ &= v_1\ (1,0, \ldots, 0) + v_2\ (0,1, \ldots 0) + \cdots + v_n\ (0,0, \ldots, 1) \\ &= v_1e_1 + v_2e_2 + \cdots + v_ne_n \\ &= \sum_{j=1}^{n} v_j e_j \end{aligned}$$

Now, $f(e_j) = f(0,0, \ldots, 1, \ldots 0) = a_1.0 + a_2.0 + \cdots + a_j.1 + \cdots + a_n.0$

Therefore, $f(e_j) = a_j$ for $j = 1,2, \ldots, n$

Thus, we can write,

$f(v) = \sum_{j=1}^{n} f(e_j)v_j$

2. Let F be a field and n be a positive integer. Consider the vector space $M_{n\times n}(F)$ consisting of all $n \times n$ matrices whose entries are from F.

Let $A = (a_{ij})_{n\times n} \in M_{n\times n}(F)$. Define

$$tr\ A = a_{11} + a_{22} + \cdots + a_{nn}$$

Then trace function is a linear functional on $M_{n\times n}(F)$ as for $A, B \in M_{n\times n}(F)$ and for $c \in F$, we have,

$$tr\ (cA + B) = \sum_{i=1}^{n}(ca_{ii} + b_{ii}) = c\sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n} b_{ii} = c\ tr\ A + trB$$

Let V^* denote the set of all linear functionals on a given vector space over a field F, that is,

$$V^* = \{f \mid f: V \to F \text{ is linear}\}.$$

Define addition and scalar multiplication on V^* as follows

$$(f + g)(v) = f(v) + g(v)$$

$$(af)(v) = a(f(v))$$

for all $v \in V$ and for all $a \in F$.

We shall show that V^* is a vector space over F.

(1) Let $f, g \in V^*$ then $f + g$ is linear as for $v_1, v_2 \in V, c \in F$, we have,

$$\begin{aligned}(f + g)(cv_1 + v_2) &= f(cv_1 + v_2) + g(cv_1 + v_2)\\ &= cf(v_1) + f(v_2) + cg(v_1) + g(v_2)\\ &= c\,(f + g)(v_1) + (f + g)(v_2)\end{aligned}$$

Therefore, $f, g \in V^* \Rightarrow f + g \in V^*$.

(2) If $f, g \in V^*$ then for any $v \in V$, we have,

$$(f + g)(v) = f(v) + g(v) = g(v) + f(v) = (g + f)(v)$$

So, $f + g = g + f,\ \forall f, g \in V^*$.

(3) Clearly, addition is associative.

(4) Define $0_V : V \to F$ by $0_V : V \to F$ by $0_V(v) = 0, \forall v \in V$. We first show that 0_V is linear. For any $v, w \in V$ and for any $c \in F$, we have,

$$0_V(cv + w) = 0 = c.0 + 0 = c\,0_V(v) + 0_V(w)$$

Therefore, $0_V \in V^*$.

Now, $(f + 0_V)(v) = f(v) + 0_V(v) = f(v) + 0 = f(v),\ \forall v \in V$

Then $f + 0_V = f, \forall f \in V^*$, that is, 0_V is the identity element in V^* under addition.

(5) For any $f \in V^*$, $(-1)f \in V^*$ and for any $v \in V$,

$$[f + (-1)f](v) = f(v) + (-1)f(v) = f(v) - f(v) = 0 = 0_V(v)$$

So, $(-1)f = -f$, that is, inverse of f under addition exists in V^* for any $f \in V^*$.

Thus, V^* is an abelian group under addition.

(6) Now, for $f, g \in V^*$ and for $a \in F$, we have, for any $v \in V$,

$$\begin{aligned}[a(f + g)](v) &= a[(f + g)(v)] = a[f(v) + g(v)] = af(v) + ag(v)\\ &= (af + ag)(v)\end{aligned}$$

Hence, $a(f + g) = af + ag$.

(7) For $a, b \in F$ and $f \in V^*$, we have, for any $v \in V$,

$$[(a + b)f](v) = (a + b)f(v) = af(v) + bf(v) = (af + bf)(v)$$

Therefore, $(a + b)f = af + bf,\ \forall f \in V^*$ and for all $a, b \in F$.

(8) For $a, b \in F$ and for $f \in V^*$, we have, for any $v \in V$,

$$[(ab)f](v) = ab\ f(v) = a[bf(v)] = a(bf)(v)$$

Therefore, $(ab)f = a(bf)$ for all $f \in V^*$ and for all $a, b \in F$.

(9) For $f \in V^*$ and $1 \in F$, we have, for any $v \in V$,

$$(1f)(v) = 1f(v) = f(v)$$

Thus, $1f = f$ for all $f \in V^*$.

Hence, V^* is a vector space over the field F.

This vector space V^* is known as **dual space of V.**

Hence dual space of V is the vector space consisting of all linear functionals on V with the operations of addition and multiplications as defined above. Now, we introduce a new concept.

Let V be a finite dimensional vector space and let $B = \{x_1, x_2, \ldots, x_n\}$ be an ordered basis for V. Let $x \in V$. Then there exist scalars $a_1, a_2, \ldots, a_n$ such that

$$x = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

For each $i = 1, 2, \ldots, n$ define $f_i : V \to F$ by $f_i(x) = a_i$.

Then f_i is called the **ith coordinate function with respect to the basis B.** Note that $f_i(x_j) = 0$ if $i \neq j$ and $f_i(x_j) = 1$ if $i = j$. In brief, we can say $f_i(x_j) = \delta_{ij}$ where

$$\delta_{ij} = \begin{cases}1 \text{ if } i = j\\ 0 \text{ if } i \neq j\end{cases}$$

δ_{ij} is known as **Kronecker delta.**

Since, $f_1, f_2, \ldots, f_n \in V^*$ and V^* is a vector space, we see that $a_1 f_1 + a_2 f_2 + \cdots + a_n f_n \in V^*$ as $a_1, a_2, \ldots, a_n$ are scalars.

Therefore, for n given scalars $c_1, c_2, \ldots, c_n$, the linear combination $f = c_1 f_1 + c_2 f_2 + \cdots + c_n f_n$ is a linear function and its value at any basis vector x_i is given by

$$f(x_i) = c_1 f_1(x_i) + c_2 f_2(x_i) + \cdots + c_i f_i(x_i) + \cdots + c_n f_n(x_i)$$

$$= c_i [\, as\ f_i(x_j) = 0\ if\ i \neq j\ and\ f_i(x_i) = 1\,]$$

Now, we wish to show that the vectors $f_1, f_2, \ldots, f_n$ are linearly independent in V^*.

If $f = c_1 f_1 + c_2 f_2 + \cdots + c_n f_n = 0_V$ where $0_V \in V^*$ be such that $0_V(x) = 0, \forall x \in V$, then $f(x_i) = 0_V(x_i) = 0$ for all $i = 1,2, \ldots, n$.

Thus, $f(x_i) = c_i = 0$ for $i = 1,2, \ldots, n$.

Therefore, the set $\{f_1, f_2, \ldots, f_n\}$ is linearly independent in V^*.

Let us define a map $T : V \to V^*$ by

$$T(x) = a_1 f_1 + a_2 f_2 + \cdots + a_n f_n = \sum_{i=1}^{n} a_i f_i$$

whence $x = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$.

It is easy to verify that T is an isomorphism and hence, V is isomorphic to V^*, the isomorphism, however, depends upon the choice of the basis in V.

Therefore, **$\dim V^* = \dim V$** if V is finite dimensional.

Thus, if $\dim V^* = \dim V = n$ then the linearly independent set $\{f_1, f_2, \ldots, f_n\}$ of vectors in V^* is a basis of V^* as any linearly independent set in V^* containing n elements is a basis of V^*.

This basis $\{f_1, f_2, \ldots, f_n\}$ is called the basisof V^*dual to the given basis $\{x_1, x_2, \ldots, x_n\}$ of V.

2.10.3 Definition : If $B = \{x_1, x_2, \ldots, x_n\}$ is an ordered basis of a vector space V and V^* be the dual space of V, then the ordered basis $B^* = \{f_1, f_2, \ldots, f_n\}$ of V^* that satisfies $f_i(x_j) = \delta_{ij}\ (1 \leq i, j \leq n)$ is called the **dual basis** of B.

2.10.4 Example. Let $V = \mathbb{R}^3$ and let $B = \{(1,0,1), (1,2,1), (0,0,1)\}$ be a basis of V. Let us try to find a basis for the dual space V^* dual to the basis B.

Let $(x, y, z) \in \mathbb{R}^3$. Since B is a basis of $\mathbb{R}^3$, there exist scalars c_1, c_2, c_3 such that

$$(x, y, z) = c_1(1,0,1) + c_2(1,2,1) + c_3(0,0,1)$$

$$= (c_1 + c_2, 2c_2, c_1 + c_2 + c_3)$$

Thus, we have, $c_1 + c_2 = x,\ 2c_2 = y,\ c_1 + c_2 + c_3 = z$. Therefore,

Solving, we get, $c_1 = x - \frac{y}{2},\ c_2 = \frac{y}{2},\ c_3 = z - x$.

$$(x, y, z) = \left(x - \frac{y}{2}\right)(1,0,1) + \frac{y}{2}(1,2,1) + (z - x)(0,0,1)$$

Hence, the dual basis $B^* = \{f_1, f_2, f_3\}$ of B is given by

$$f_1(x, y, z) = x - \frac{y}{2}$$

$$f_2(x, y, z) = \frac{y}{2}$$

$$f_3(x, y, z) = z - x \quad \blacksquare$$

2.10.5 Theorem. **Let V be a finite dimensional vector space over a field F with the ordered basis $B = \{x_1, x_2, \ldots, x_n\}$. Let $B^* = \{f_1, f_2, \ldots, f_n\}$ be the ordered basis for V^* dual to B. Then for any $f \in V^*$, we have,**

$$f = \sum_{i=1}^{n} f(x_i) f_i.$$

Proof. Since $f(x_i) \in F$ for $i = 1,2, \ldots, n$ and B^* generates V^*, let

$$g = \sum_{i=1}^{n} f(x_i) f_i$$

Now, for $1 \leq k \leq n$, we have,

$$g(x_k) = \sum_{i=1}^{n} f(x_i) f_i(x_k)$$

$$= f(x_1) f_1(x_k) + f(x_2) f_2(x_k) + \cdots + f(x_k) f_k(x_k) + \cdots + f(x_n) f_n(x_k)$$

$$= f(x_k)[\, as\ f_i(x_k) = 0\ for\ i \neq k,\ f_k(x_k) = 1]$$

Thus, $g(x_k) = f(x_k)$ for all basis vectors x_k.

Let $x \in V$. Then there exist scalars $c_1, c_2, \ldots, c_n$ such that

$$x = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

Since, g is linear, we have,

$$g(x) = c_1 g(x_1) + c_2 g(x_2) + \cdots + c_n g(x_n)$$

$= c_1 f(x_1) + c_2 f(x_2) + \cdots + c_n f(x_n)[g(x_k) = f(x_k), k = 1,2,\ldots,n]$

$= f(c_1 x_1 + c_2 x_2 + \cdots + c_n x_n)[\, f \text{ is linear}]$

$= f(x)$

Therefore, $g = f$. Hence,

$$f = \sum_{i=1}^{n} f(x_i) f_i \blacksquare$$

2.11 THE DOUBLE DUAL

In the last section, we have seen that for a given ordered basis B for a vector space V, there is an ordered basis for V^* dual of B. But this leads us to a problem *'Is every ordered basis for V^* is dual of some ordered basis for V?'* We will seek answer to this problem in this section.

Let us consider the dual space of V^* as V^{**}.

Each vector $x \in V$ induces a linear functional L_x on V^*[i.e. $L_x : V^* \to F$] defined by

$$L_x(f) = f(x), \ \forall f \in V^*.$$

We first show that L_x is linear. In fact, for $f, g \in V^*$ and for $c \in F$, we have,

$$L_x(cf + g) = (cf + g)(x) = (cf)(x) + g(x)$$

$$= c\, f(x) + g(x) = cL_x(f) + L_x(g)$$

Thus, $L_x \in V^{**}$.

2.11.1 Theorem. Let V be a finite dimensional vector space, and define $\psi : V \to V^{**}$ by $\psi(x) = L_x$.
Then ψ is an isomorphism. In other words, V is isomorphic to V^{**}.

Proof. To prove this theorem, we prove this lemma first.

Lemma : Let V be a finite dimensional vector space, and let $x \in V$. If $L_x(f) = 0$ for all $f \in V^*$, then $x = 0$.

Proof of Lemma : Let $x \neq 0$. We show that there exists $f \in V^*$ such that $L_x(f) \neq 0$. Let us choose an ordered basis $B = \{x_1, x_2, \ldots, x_n\}$ for V such that $x_1 = x$. Let $\{f_1, f_2, \ldots, f_n\}$ be the dual basis of B. Then $f_1(x_1) = 1$.

So, $f_1(x) = 1` \neq 0$ $(as\ x = x_1)$. Thus, $f = f_1$ serves our purpose. Thus, $x = 0$.

Proof of the theorem : We first show that ψ is linear. Let $x, y \in V$ and $c \in F$. Then for $f \in V^*$, we have,

$$\psi(cx + y)(f) = L_{cx+y}(f) = f(cx + y) = cf(x) + f(y) = cL_x(f) + L_y(f)$$

$$= (cL_x + L_y)(f)$$

Thus, $\psi(cx + y) = cL_x + L_y = c\psi(x) + \psi(y)$

Hence, ψ is linear.

Let $x \in \ker\psi$. Then $\psi(x)$ is the zero functional on V^* for $x \in V$. Then $L_x(f) = 0$ for all $f \in V^*$. Then by the lemma, we have, $x = 0$.

Therefore, $\ker\psi = \{0\}$.

So, ψ is one-one.

Clearly, ψ is onto as ψ is one-one and $\dim V = \dim V^* = \dim V^{**}$.

Hence, ψ is an isomorphism. In other words, $V \simeq V^{**} \blacksquare$

Corollary : Let V be a finite dimensional vector space with dual space V^*. Then every ordered basis for V^* is the dual basis for some basis for V.

Proof. Let $\{f_1, f_2, \ldots, f_n\}$ be an ordered basis for V^*. Then there exists a dual basis $\{L_{x_1}, L_{x_2}, \ldots, L_{x_n}\}$ in V^{**}, that is, $\delta_{ij} = L_{x_i}(f_j) = f_j(x_i)$ for all i and j. Thus, $\{f_1, f_2, \ldots, f_n\}$ is the dual basis of $\{x_1, x_2, \ldots, x_n\}$.

2.11.2 Definition. The vector space V^{**} is called the **double dual of V.**

2.12 TRANSPOSE OF A LINEAR TRANSFORMATION AND ITS MATRIX IN THE DUAL BASIS

Let us consider two finite dimensional vector spaces V and W over the same field F and a linear transformation T from V into W. Suppose $g : W \to F$ be linear, that is, g is a linear functional on W, and let $f(v) = g(T(v))$ for each v in V. Thus, $f = gT$, that is, $f : V \to F$, a function, is the composition of T ($T : V \to W$) with g ($g : W \to F$). Since g, T both are linear and composition of two linear functions is also linear, we see that f is also linear. Hence, f is a linear functional on V. If we write $f = T^t g$ then we have, $T^t g(v) = g(T(v))$ for all $v \in V$. Thus, T induces a mapping T^t from W^*(as $g \in W^*$) into V^* (as $T^t g = f \in V^*$). We now show that, T^t is linear. In fact, for $g_1, g_2 \in W^*$ and for $c \in F$,

$$[T^t(cg_1 + g_2)](v) = (cg_1 + g_2)(T(v))$$

$$= cg_1(T(v)) + g_2(T(v))$$

$$= c(T^t g_1)(v) + (T^t g_2)(v)$$

$$= (cT^t g_1 + g_2)(v)$$

Thus, $T^t(cg_1 + g_2) = cT^t g_1 + g_2$

So, T^t is linear.

We shall call T^t the transpose of T. Some authors prefer to call it adjoint (or Dual) of T.

2.12.1 Properties :

(i) $0^t = 0$

(ii) $I^t = I$

(iii) $(A+B)^t = A^t + B^t$

(iv) $(AB)^t = B^t A^t$

(v) $(A^{-1})^t = (A^t)^{-1}$

Proof. Let V and W be two finite dimensional vector spaces over the same field F and let $V^*$and W^*be the dual spaces$V$ and W respectively.

(i) $0^t(g)(v) = g(0(v)) = g(0) = 0 = 0(g(v))$, for all $g \in W^*$ and for all $v \in V$.So, $0^t = 0$.

(ii) Exercise.

(iii) Let $A, B : V \to W$ be two linear maps. Then $A+B : V \to W$ defined by $(A+B)(v) = A(v) + B(v)$is also linear. Hence, $A^t, B^t, (A+B)^t, A^t + B^t$ all are linear maps from W^* into V^*. For any $g \in W^*$ and for any $v \in V$

$$\begin{aligned}(A+B)^t(g)(v) &= g[(A+B)](v)\\ &= g[(A+B)(v)]\\ &= g[A(v) + B(v)]\\ &= g(A)(v) + g(B)(v)\ [\,as\ g\ is\ linear]\\ &= A^t(g)(v) + B^t(g)(v)\\ &= (A^t + B^t)(g)(v)\end{aligned}$$

Hence, $(A+B)^t = A^t + B^t$.

(iv) Clearly,

$(AB)^t(g)(v) = g(AB)(v) = (gA)(B)(v) = B^t(gA)(v) = B^t A^t(g)(v)$

Thus, $(AB)^t = B^t A^t$.

(v) Since, A is invertible, we have, $AA^{-1} = A^{-1}A = I$

Using property (ii) and property (iv), we have,

$$(A^{-1})^t A^t = A^t (A^{-1})^t = I^t = I.$$

Thus, $(A^{-1})^t = (A^t)^{-1}$.

2.12.2Theorem. Let V and W be finite dimensional vector spaces over the field F. Let β be an ordered basis for V with dual basis β^*, and let β' be an ordered basis for W with dual basis β'^*. Let T be a linear transformation from V into W; let $A = (a_{ij})$ be the matrix of T relative to β, β' and let$B = (b_{ij})$ be the matrix of T^t relative to β'^*, β^*. Then $b_{ij} = a_{ji}$.

Proof. Let
$$\beta = \{v_1, v_2, .., v_n\},\ \beta' = \{w_1, w_2, .., w_m\}$$
$$\beta^* = \{f_1, f_2, ..., f_n\},\ \beta'^* = \{g_1, g_2, ..., g_m\}.$$

We have,

$$T(v_i) = \sum_{k=1}^{m} a_{ki} w_k\ ,\quad i = 1,2,..,n$$

$$T^t(g_j) = \sum_{i=1}^{n} b_{ij} f_i\ ,\quad j = 1,2,..,m.$$

Again,

$$\begin{aligned}(T^t g_j)(v_i) &= g_j(Tv_i)\\ &= g_j\left(\sum_{k=1}^{m} a_{ki} w_k\right)\\ &= \sum_{k=1}^{m} a_{ki}\, g_j(w_k)\ [g_j\ linear]\\ &= \sum_{k=1}^{m} a_{ki}\, \delta_{jk}\\ &= a_{ji}\,[\delta_{jk} = 0\ for\ j \neq k, \delta_{jj} = 1]\end{aligned}$$

For any linear functional f on V

$$f = \sum_{i=1}^{n} f(v_i) f_i\,.$$

Taking $f = T^t g_j$ and using the fact $(T^t g)(v_i) = a_{ji}$, we have,

$$\sum_{i=}^{n} b_{ij} f_i = T^t g_j = \sum_{i=1}^{n} [(T^t g_j)(v_i)] f_i = \sum_{i=1}^{n} a_{ji} f_i$$

Hence, by linear independence of $\{f_1, f_2, ..., f_n\}$, we have $b_{ij} = a_{ji}$ ■

Thus, we can say that, if A be the matrix representation of a linear transformation $T : V \to W$ with respect to some pair of ordered bases then A^t be the matrix representation of $T^t : W^* \to V^*$ with respect to the pair of bases dual to the former pair.

2.12.3 Definition. If V is a vector space over the field F and S is a subset of V, then S^0, the **annihilator** of S, is defined as

$$S^0 = \{ f \in V^* : \ f(s) = 0 \ \forall s \in S\}.$$

That is, S^0 is the set of all linear functionals f on V such that $f(s) = 0$, for all $s \in S$. In other words, S^0 is the collection of all those linear functiponals on V which take every element of S to 0. Clearly, 0_f, the zero functional on V, belongs to S^0 for any $S \subseteq V$. It is easy to see that S^0 is a subspace of V^* whether S is a subspace of V or not. In fact, $0_f \in S^0$ and for $f, g \in S^0$ and for $c \in F$, we have,

$$(cf + g)(s) = cf(s) + g(s) = 0, \ \forall s \in S$$

Hence, $f, g \in S^0$ and $c \in F$ implies $cf + g \in S^0$. So, S^0 is a subspace of V^*.

If $S = \{0_V\}$, that is, S is the set consisting of zero vector only, then $S^0 = V^*$. If $S = V$, then S^0 is the zero subspace of V^*.

2.12.4 Proposition . Let S, T be two subsets of a vector space over a field F. Then

(i) $S \subseteq T$ implies $T^0 \subseteq S^0$.

(ii) If $W = L(S)$ then $W^0 = S^0$.

Proof (i) Let $f \in T^0$. Now, $s \in S$ implies $s \in T$ $(as\ S \subseteq T)$. Thus, $f(s) = 0$.

Therefore, $f(s) = 0$ for all $s \in S$. So, $f \in S^0$.

Hence, $T^0 \subseteq S^0$.

(ii) Since, $S \subseteq W$, by (i), $W^0 \subseteq S^0$.

Let $f \in S^0$ and let $w \in W$. Then $f(s) = 0$ for all $s \in S$. Now, $w = \Sigma c_i s_i$ where $c_i \in F, s_i \in S \forall i$.

Thus, $f(w) = \Sigma c_i f(s_i) = 0$. Therefore, $f \in W^0$.

Thus, $S^0 \subseteq W^0$.

Hence, $W^0 = S^0$.

Now, we shall try to prove an important theorem.

2.12.5 Theorem. Let V be a finite dimensional vector space over the field F, and let W be a subspace of V. Then

$$\boldsymbol{dim\, W + dim\, W^0 = dim\, V.}$$

Proof. Let $\dim V = n$ and $\dim W = k$. Let $\{v_1, v_2, .., v_k\}$ be a basis for W. Let us extend the basis $\{v_1, v_2, .., v_k\}$ for W to the basis for V as $B = \{v_1, v_2, ..., v_k, v_{k+1}, ..., v_n\}$. Let $\{f_1, f_2, ... f_k, f_{k+1}, ..., f_n\}$ be the basis for V^* which is dual to this basis B for V.

We shall show that $\{f_{k+1}, .., f_n\}$ is a basis for the annihilator W^0.

Clearly,
$$f_i(v_j) = \delta_{ij} = \begin{cases} 0 \ if \ i \neq j \\ 1 \ if \ i = j \end{cases}$$

Thus, $f_i(v_j) = 0$ if $i \in \{k+1, k+2, ..., n\}$ and $j \in \{1,2,..,k\}$.

If $w \in W$, then $w = c_1v_1 + c_2v_2 + \cdots + c_kv_k$ for some scalars $c_1, c_2, ..., c_k$.

Then for $i \in \{k+1, k+2, ..., n\}$, we have,

$$f_i(w) = c_1 f_i(v_1) + c_2 f_i(v_2) + \cdots + c_k f_i(v_k) = 0.$$

Thus, $f_i \in W^0$ for $i = k+1, k+2, ..., n$.

Since $\{f_{k+1}, f_{k+2}, ..., f_n\}$ is independent in W^0, it is enough to show that $\{f_{k+1}, f_{k+2}, ..., f_n\}$ spans W^0.

Let $f \in V^*$. Now,

$$f = \sum_{i=1}^{n} f(v_i) f_i$$

Now, if $f \in W^0$, then $f(v_1) = f(v_2) = \cdots = f(v_k) = 0$, so

$$f = \sum_{i=k+1}^{n} f(v_i) f_i$$

Hence, $f \in L\{f_{k+1}, f_{k+2}, ..., f_n\}$. In other words, $W^0 = L\{f_{k+1}, f_{k+2}, ..., f_n\}$.

Thus, $\{f_{k+1}, f_{k+2}, ..., f_n\}$ is a basis for W^0. Therefore, $\dim W^0 = n - k$. Thus,

$$k + (n-k) = n \Rightarrow \dim W + \dim W^0 = \dim V \blacksquare$$

Corollary : If S is any subset of a finite dimensional vector space V, then $(S^0)^0$ is the subspace spanned by S.

Proof. Let $W = L(S)$. Then by proposition 2.12.4, we have, $W^0 = S^0$.

So, it is enough to proof, $S^{00} = W^{00} = W$.

By the theorem, $\dim W + \dim W^0 = \dim V$.

Thus, $\dim W^0 + \dim W^{00} = \dim V^* = \dim V$.

In other words, $\dim W = \dim W^{00}$.

Since W is a subspace of W^{00}, we have, $W = W^{00}$

2.12.6 Example. Let $V = \mathbb{R}^3$, $S = \{(1,2,-1),(3,0,1)\}$

For any $f \in V^*$, we have,

$$f(x,y,z) = ax + by + cz, \qquad a,b,c \in F$$

Now, $f \in S^0$ if and only if $f(1,2,-1) = 0$ and $f(3,0,1) = 0$. Thus,

$$f \in S^0 <=> \begin{array}{l} a + 2b - c = 0 \\ 3a + 0b + c = 0 \end{array}$$

$$<=> \frac{a}{2} = \frac{b}{-4} = \frac{c}{-6}$$

$$<=> (a,b,c) = (s,-2s,-3s), s \in F$$

Thus, $f \in S^0$ if and only if

$$f(x,y,z) = sx - 2sy - 3sz = s(x - 2y - 3z) = sg(x,y,z)$$

where $g(x,y,z) = x - 2y - 3z$.

Hence, $f = sg, s \in F$.

Therefore, $S^0 = L\{g\}$ and hence, $\dim S^0 = 1$.

Now, it is easy to see that S is linearly independent and if $W = L(S)$ then W is a subspace of V such that $\dim W = 2$.

Again, by proposition 2.12.4, $\dim W^0 = \dim S^0 = 1$.

Hence, $\dim W + \dim W^0 = 2 + 1 = 3 = \dim V$ ∎

2.12.7 Theorem. Let V and W be vector spaces over the same field F, and let T be a linear transformation from V into W. The null space of T^t is the annihilator of the range of T. If V and W are finite dimensional, then

(i) $rank\ (T^t) = rank\ (T)$

(ii) the range of T^t is the annihilator of the null space of T.

Proof. Let $g \in W^*$, then we have,

$$(T^t g)(v) = g(T(v)), \quad \forall v \in V.$$

Now, $g \in null\ T^t <=> (T^t g)(v) = g(T(v)) = 0, \forall v \in V$

$$<=> g \in annihilator\ of\ the\ range\ T.$$

Hence, the null space of T^t is the annihilator of the range of T.

Let V and W are two finite dimensional vector spaces over the field F where $\dim V = n$ and $\dim W = m$. Let us denote range of a linear functional f by R_f and null space of f by N_f.

(i) Let $rank\ T = r$. Then $r = \dim R_T$. Thus, we have,

$$\dim R_T + \dim R_T^0 = \dim W = m.$$

Therefore, $\dim R_T^0 = m - r$.

So, by the first part of the theorem, $\dim N_T = \dim R_T^0 = m - r$.

Again, since, $T^t : W^* \to V^*$ is linear, we have,

$$\dim N_{T^t} + \dim R_{T^t} = \dim W^* = \dim W = m$$

Therefore, $\dim R_{T^t} = m - (m - r) = r$.

Hence, $rank\ (T^t) = rank\ T$.

(ii) Let $f \in R_{T^t}$. Then there exists $g \in W^*$ such that $f = T^t g$.

Now, if $v \in N_T$ then $T(v) = 0$.

Therefore, $f(v) = (T^t g)(v) = g(T(v)) = g(0) = 0$.

Thus, $f \in N_T^0$. That is, $R_{T^t} \subseteq N_T^0$.

In fact, R_{T^t} is a subspace of N_T^0.

Now, $\dim N_T + \dim N_T^0 = \dim V = n$

gives $\dim N_T^0 = n - \dim N_T$.

Again,

$\dim N_T + \dim R_T = \dim V = n$ gives $\dim R_{T^t} = \dim R_T = n - \dim N_T$.

Hence, $\dim N_T^0 = \dim R_{T^t}$.

Thus, $N_T^0 = R_{T^t}$, that is, the range of T^t equals the null space of T.

Solved Examples :

1. In $\mathbb{R}^3$, let $v_1 = (1,0,1), v_2 = (0,1,-2), v_3 = (-1,-1,0)$,
   (a) If f is a linear functional on $\mathbb{R}^3$ such that $f(v_1) = 1$, $f(v_2) = -1, f(v_3) = 3$ and if $v = (a,b,c)$, find $f(v)$.
   (b) Describe explicitly a linear functional f on $\mathbb{R}^3$ such that $f(v_1) = f(v_2) = 0$ but $f(v_3) \neq 0$.
   (c) Let f be any linear functional such that $f(v_1) = f(v_2) = 0$ and $f(v_3) \neq 0$. If $v = (2,3,-1)$, show that $f(v) \neq 0$.

Solution.(a) Since,

$$\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ -1 & -1 & 0 \end{vmatrix} \neq 0$$

we see that, $\{v_1, v_2, v_3\}$ is a basis for $\mathbb{R}^3$.

Let $v = (a,b,c) \in \mathbb{R}^3$. Then there exist scalars p,q,r such that

$$v = pv_1 + qv_2 + rv_3$$

i.e. $(a,b,c) = p(1,0,1) + q(0,1,-2) + r(-1,-1,0)$

i.e. $p - r = a,\ q - r = b,\ p - 2q = c$

Solving we get,

$p = 2a - 2b - c, q = a - b - c$ and $r = a - 2b - c.$

Now, we have,

$$f(v) = f(a,b,c) = pf(v_1) + qf(v_2) + rf(v_3)[\text{ as } f \text{ is linear}]$$
$$= (2a - 2b - c).1 + (a - b - c).(-1) + (a - 2b - c).3$$
$$= 4a - 7b - 3c$$

(b) Let $f(x,y,z) = x - 2y - z$. Then

$$f(v_1) = f(1,0,1) = 0, f(v_2) = f(0,1,-2) = 0,$$
$$f(v_3) = f(-1,-1,0) = 1 \neq 0.$$

(c) If $v = (2,3,1)$, we have by (a)

$$(2,3,-1) = -v_1 - 3v_3 \text{ [taking } a = 2. b = 3, c = -1]$$

Thus, $f(2,3,-1) = -f(v_1) - 3f(v_3) \neq 0$ as $f(v_1) = 0$ but $f(v_3) \neq 0$.

2. **If $B = \{v_1, v_2, v_3\}$ be an ordered basis for $\mathbb{C}^3$ defined by $v_1 = (1,0,-1)$, $v_2 = (1,1,1), v_3 = (2,2,0)$. Find the dual basis of B.**

Solution. Let $(x,y,z) \in \mathbb{C}^3$ and a,b,c be the scalars such that

$$(x,y,z) = a(1,0,-1) + b(1,1,1) + c(2,2,0)$$
$$= (a + b + 2c, b + 2c, -a + b)$$

Thus,

$$a + b + 2c = x$$
$$b + 2c = y$$
$$-a + b \quad = z$$

Solving, $a = x - y, b = x - y + z, c = -\frac{1}{2}x + y - \frac{1}{2}z$. Therefore,

$$(x,y,z) = (x - y)(1,0,-1) + (x - y + z)(1,1,1) + \left(-\frac{1}{2}x + y - \frac{1}{2}z\right)(2,2,0)$$

Thus, if $\{f_1, f_2, f_3\}$ be the dual basis of B, then f_1, f_2, f_3 are given by

$$f_1(x,y,z) = x - y$$
$$f_2(x,y,z) = x - y + z$$
$$f_3(x,y,z) = -\frac{1}{2}x + y - \frac{1}{2}z \blacksquare$$

3. **Let m and n be positive integers and F be a field. Let $f_1, f_2, \dots, f_m$ be linear functional on F^n. For $\alpha \in F^n$, define, $T(\alpha) = (f_1(\alpha), f_2(\alpha), \dots, f_m(\alpha))$. Show that T is a linear transformation from F^n to F^m.**

Solution. Clearly, T is well defined. Now, for $\alpha, \beta \in F^n$ and for $c \in F$, we have,

$$T(c\alpha + \beta) = (f_1(c\alpha + \beta), f_2(c\alpha + \beta), \dots, f_m(c\alpha + \beta))$$
$$= (cf_1(\alpha) + f_1(\beta), cf_2(\alpha) + f_2(\beta), \dots, cf_m(\alpha) + f_m(\beta))$$

[as f_i's are linear]

$$= (cf_1(\alpha), cf_2(\alpha), \dots, cf_m(\alpha)) + (f_1(\beta), f_2(\beta) \dots, f_m(\beta))$$
$$= c(f_1(\alpha), f_2(\alpha), \dots, f_m(\alpha)) + (f_1(\beta), f_2(\beta) \dots, f_m(\beta))$$
$$= c\,T(\alpha) + T(\beta)$$

Hence, T is linear.

4. **Define $f \in (\mathbb{R}^2)^*$ by $f(x,y) = 2x + y$ and $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x,y) = (3x + 2y, x)$.**
   (a) Compute $T^t(f)$
   (b) Compute $[T^t]_{\beta^*}$, that is, the matrix represented by T^t with respect to the ordered basis β^*, where β is the standard ordered basis for $\mathbb{R}^2$ and $\beta^* = \{f_1, f_2\}$ is the dual basis, by finding scalars a,b,c,d such that $T^t(f_1) = af_1 + cf_2$ and $T^t(f_2) = bf_1 + df_2$.
   (c) Compute $[T]_\beta$ and $([T]_\beta)^t$ and compare your results with (b).

Solution. (a) We have, for $(x,y) \in \mathbb{R}^2$,

$$T^t(f)(x,y) = f(T(x,y)) = f(3x + 2y, x) = 2(3x + 2y) + x = 7x + 4y.$$

(b) Let $B = \{(1,0),(0,1)\}$ be the standard basis for $\mathbb{R}^2$. Thus, for $(x,y) \in \mathbb{R}^2$, we have,

$$(x,y) = x(1,0) + y(0,1).$$

Therefore, $f_1(x,y) = x$ and $f_2(x,y) = y$. Now,

$$T^t(f_1)(x,y) = f_1(T(x,y)) = f_1(3x+2y,x) = 3x+2y$$

and $$= 3f_1(x,y) + 2f_2(x,y)$$

$$T^t(f_2)(x,y) = f_2(T(x,y)) = f_2(3x+2y,x) = x = 1f_1(x,y) + 0f_2(x,y).$$

Hence, we have,

$$[T^t]_{B^*} = \begin{pmatrix} 3 & 1 \\ 2 & 0 \end{pmatrix}$$

(c) Since, $T(x,y) = (3x+2y, x)$, we have,

So, $$[T]_B = \begin{pmatrix} 3 & 2 \\ 1 & 0 \end{pmatrix}$$

$$([T]_B)^t = \begin{pmatrix} 3 & 1 \\ 2 & 0 \end{pmatrix}.$$

Hence, we see that $[T^t]_{B^*} = ([T]_B)^t$.

5. **Let $v_1 = (1,0,-1,2)$ and $v_2 = (2,3,1,1)$ and let W be the subspace of $\mathbb{R}^4$ spanned by v_1 and v_2. Which linear functionals f :**

$$f(x_1,x_2,x_3,x_4) = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4$$

are in the annihilator of W ?

Solution. Clearly, v_1 and v_2 are linearly independent as neither is a scalar multiple of other. Thus dimension of $W = L\{v_1, v_2\}$ is 2.

If $f \in W^0$ then we have, $f(v_1) = 0, f(v_2) = 0$. Now,

$$f(v_1) = 0 \Rightarrow c_1 - c_3 + 2c_4 = 0 \Rightarrow c_1 = c_3 - 2c_4$$

$$f(v_2) = 0 \Rightarrow 2c_1 + 3c_2 + c_3 + c_4 = 0 \Rightarrow 2c_1 + 3c_2 = -c_3 - c_4$$

Thus, $2(c_3 - 2c_4) + 3c_2 = -c_3 - c_4 \Rightarrow c_2 = -c_3 + c_4$.

Hence, $f \in W^0$ if and only if

$$f(x_1,x_2,x_3,x_4) = (c_3 - 2c_4)x_1 + (-c_3 + c_4)x_2 + c_3x_3 + c_4x_4$$

where c_3, c_4 are arbitrary.

6. **Let W be the subspace of $\mathbb{R}^5$ which is spanned by the vectors**

$$v_1 = e_1 + 2e_2 + e_3,$$

$$v_2 = e_2 + 3e_3 + 3e_4 + e_5,$$

$$v_3 = e_1 + 4e_2 + 6e_3 + 4e_4 + e_5$$

where e_i's are standard basis of $\mathbb{R}^5$. Find W^0.

Solution. By the problem, we have,

$$v_1 = (1,2,1,0,0), \quad v_2 = (0,1,3,3,1), \quad v_3 = (1,4,6,4,1)$$

If $f \in W^0$, then $f(v_1) = 0, f(v_2) = 0, f(v_3) = 0$.

Now, every functional f can be written as

$$f(x_1,x_2,x_3,x_4,x_5) = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 + c_5x_5.$$

Thus, $$f(v_1) = 0 \Rightarrow c_1 + 2c_2 + c_3 = 0$$
$$f(v_2) = c_2 + 3c_3 + 3c_4 + c_5 = 0$$
$$f(v_3) = c_1 + 4c_2 + 6c_3 + 4c_4 + c_5 = 0$$

which gives $c_1 = -4c_4 - 3c_5$, $c_2 = 3c_4 + 2c_5$, $c_3 = -2c_4 - c_5$. Hence, we have, $f \in W^0$ if and only if,

$$f(x_1,x_2,x_3,x_4,x_5) = (-4c_4 - 3c_5)x_1 + (3c_4 + 2c_5)x_2 - (2c_4 + c_5)x_3 + c_4x_4 + c_5x_5$$

where c_4, c_5 are arbitrary.

7. **Let V be the vector space of all 2×2 matrices over the field of real numbers, and let**

$$B = \begin{pmatrix} 2 & -2 \\ -1 & 1 \end{pmatrix}.$$

Let W be the subspace of V consisting of all A such that $AB = 0$. Let f be a linear functional on V which is in the annihilator of W. Suppose that $f(I) = 0$ and $f(C) = 3$, where I is the 2×2 identity matrix and

$$C = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Find $f(B)$.

Solution. We know that if f is a linear functional on V and $A = (a_{ij})_{2\times 2} \in V$ then

$$f(A) = aa_{11} + ba_{12} + ca_{21} + da_{22} \text{ where } a,b,c,d \in \mathbb{R}.$$

If $A = \begin{pmatrix} x & y \\ z & w \end{pmatrix} \in W$, then $AB = 0$ gives

$$\begin{pmatrix} x & y \\ z & w \end{pmatrix}\begin{pmatrix} 2 & -2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Rightarrow 2x - y = 0, \quad -2z + w = 0 \Rightarrow y = 2x, w = 2z$$

Thus, $A \in W \Rightarrow A = \begin{pmatrix} x & 2x \\ z & 2z \end{pmatrix}$. Now, $f \in W^0 \Rightarrow f(A) = 0$. Therefore,

$$f(A) = ax + 2bx + cz + 2dz = 0, \text{ i.e. } (a+2b)x + (c+2d)z = 0. \quad \forall x, z \in \mathbb{R}$$

It happens only when $a + 2b = 0$ and $c + 2d = 0$ i.e. $b = -\frac{a}{2}, d = -\frac{c}{2}$.

Thus, $f(A) = aa_{11} - \frac{1}{2}aa_{12} + ca_{21} - \frac{1}{2}ca_{22}$.

Given that $f(I) = 0$ i.e. $f\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix} = 0$, which shows that $a - \frac{1}{2}c = 0$ i.e.$c = 2a$.

and $f(C) = 0$ i.e. $f\begin{pmatrix}0 & 0\\0 & 1\end{pmatrix} = 3 \Rightarrow -\frac{1}{2}c = 3 \Rightarrow c = -6$. So, $a = -3$. Hence,

$$f(B) = f\begin{pmatrix}2 & -2\\-1 & 1\end{pmatrix} = 2a + \left(-\frac{1}{2}a\right)(-2) + c(-1) + \left(-\frac{1}{2}c\right).1$$
$$= 2(-3) + (-3) + (-6)(-1) + 3.1$$
$$= 0$$

Hence, $f(B) = 0$ ■

8. **Let W_1 and W_2 be two subspaces of a finite dimensional vector space V. Prove that $(W_1 + W_2)^0 = W_1^0 \cap W_2^0$**

Solution. Let $f \in (W_1 + W_2)^0$. Then $f(w) = 0$ for all $w \in W_1 + W_2$.

Let $w_1 \in W_1$. Since, $0 \in W_2$. We have, $w_1 = w_1 + 0 \in W_1 + W_2$. Then $f(w_1) = 0$. So, $f \in W_1^0$.

Similarly, it can be shown that $f \in W_2^0$. Thus, $f \in W_1^0 \cap W_2^0$.

Therefore, $(W_1 + W_2)^0 \subseteq W_1^0 \cap W_2^0$.

Conversely, let $g \in W_1^0 \cap W_2^0$.

Then $g(w_1) = 0, g(w_2) = 0$ for all $w_1 \in W_1, w_2 \in W_2$.

Now, $w \in W_1 + W_2$. Then $w = w_1 + w_2$ where $w_1 \in W_1, w_2 \in W_2$.

Then $g(w) = g(w_1 + w_2) = g(w_1) + g(w_2) = 0 + 0 = 0$.

Therefore, $g \in (W_1 + W_2)^0$.

So, $W_1^0 \cap W_2^0 \subseteq (W_1 + W_2)^0$.

Hence, $(W_1 + W_2)^0 = W_1^0 \cap W_2^0$.

9. **Suppose that W is a finite dimensional vector space and that $T: V \to W$ is linear. Prove that $N(T^t) = (R(T))^0$ where $N(T^t)$ is the null space of T^t and $R(T)$ be the range of T.**

Solution. Let $f \in N(T^t)$. Then $T^t(f) = 0$. In other words, $T^t(f)(v) = 0$ for all $v \in V$. Thus,

$$f(T(v)) = 0\ \forall v \in V \Rightarrow f \in (T(V))^0 = (R(T))^0$$

Therefore, $N(T^t) \subseteq (R(T))^0$.

Again, $g(R(T))^0 \Rightarrow g(T(V)) = 0 \Rightarrow g(T(v)) = 0, \forall v \in V$
$$\Rightarrow T^t(g)(v) = 0\ \forall v \in V$$

Thus, $T^t g = 0$ which shows that $g \in N(T^t)$.

Therefore, $(R(T))^0 \subseteq N(T^t)$.

Hence, $(N(T^t)) = (R(T))^0$.

10. **Let V be a finite dimensional vector space over the field F and let W be a subspace of V. If f is a linear functional on W, prove that there exists a linear functional g on V such that $g(\alpha) = f(\alpha)$ for each $\alpha \in W$.**

Solution. Let $B = \{w_1, w_2, \dots, w_k\}$ be a basis for W. Since, B is independent in V, it can be extended to a basis $B' = \{w_1, w_2, \dots, w_k, v_{k+1}, v_{k+2}, \dots, v_n\}$ for V.

Since a linear function on a vector space is uniquely determined by its values on a basis and conversely any function on the basis can be extended to a linear function on the space, let us define h from B' to F by

$h(w_i) = f(w_i)$ for $i = 1,2, \dots k$ and $h(v_i) = 0$ for $i = k+1, k+2, \dots, n$.

Let $v \in V$. Then there exist scalars $p_1, p_2, \dots, p_k, p_{k+1}, \dots, p_n$ such that

$$v = p_1 w_1 + p_2 w_2 + \dots + p_k w_k + p_{k+1} v_{k+1} + \dots + p_n v_n$$

Since, h is defined on B', h defines a linear functional g on V by

$$g(v) = p_1 g(w_1) + \dots + p_k g(w_k) + p_{k+1} g(v_{k+1}) + \dots + p_n g(v_n)$$

Let $\alpha \in W$. Then there exist $c_1, c_2, \dots, c_k$ such that $\alpha = c_1 w_1 + c_2 w_2 + \dots + c_k w_k$. Thus,

$$f(\alpha) = f(c_1 w_1 + c_2 w_2 + \dots + c_k w_k) = c_1 f(w_1) + c_2 f(w_2) + \dots + c_k f(w_k)$$
$$= c_1 g(w_1) + c_2 g(w_2) + \dots + c_k g(w_k)$$
$$= g(c_1 w_1 + c_2 w_2 + \dots + c_k w_k)$$
$$= g(\alpha)$$

Hence, $f = g$ on W.

11. **Let F be a subfield of the field of complex numbers and let V be any vector space over F. Suppose that f and g are linear functionals on V**

such that the function h defined by $h(v) = f(v)g(v)$ is also a linear functional on V. Prove that either $f = 0$ or $g = 0$.

Solution. If possible, let $f \neq 0$ and $g \neq 0$ on V. Let $v \in V$. Then

$$h(2v) = f(2v)g(2v) = 4f(v)g(v)[\,as\ f, g\ are\ linear]$$

again, $h(2v) = 2h(v)[\,h\ is\ linear] = 2f(v)g(v)$. Therefore,

$$4f(v)g(v) = 2f(v)g(v)$$

which shows that $f(v)g(v) = 0$ for all $v \in V$. ....(A)

Let B be basis for V.

Let $B_1 = \{b \in B : f(b) = 0\}$ and $B_2 = \{b \in B : g(b) = 0\}$.

Then $B = B_1 \cup B_2$.

If $B_1 \subseteq B_2$, then $B = B_2$ and it shows that $g = 0$ on B and hence on V which is not the case.

So, $B_1 \nsubseteq B_2$. Similarly, it can be said that $B_2 \nsubseteq B_1$.

Let us choose $b_1 \in B_1 - B_2$ and $b_2 \in B_2 - B_1$.

Then $f(b_2) \neq 0$ and $g(b_1) \neq 0$. Then

$$f(b_1 + b_2)g(b_1 + b_2) = [f(b_1) + f(b_2)][g(b_1) + g(b_2)]$$
$$= f(b_1)g(b_1) + f(b_1)g(b_2) + f(b_2)g(b_1) + f(b_2)g(b_2)$$
$$= f(b_2)g(b_1)[\,as\ f(b_1) = 0 = g(b_2)] \neq 0$$

which contradicts (A).

hence, either $f = 0$ or $g = 0$ on V.

Exercise

1. **Let V be a finite dimensional vector space over the field F and let W be a subspace of V. If f is a linear functional on W, prove that there is a linear functional g on V such that $g(v) = f(v)$ for each v in the subspace W.**

2. **Let V be a nonzero vector space, and let W be a proper subspace of V (i.e. $W \neq V$). Prove that there exists a non-zero linear functional $f \in V^*$ such that $f(x) = 0$ for all $x \in W$.**

3. **Let $V = \mathbb{R}^3$ and define $f_1, f_2, f_3 \in V^*$ as follows**
   $f_1(x, y, z) = x - 2y,$
   $f_2(x, y, z) = x + y + z,$
   $f_3(x, y, z) = y - 3z.$
   Prove that $\{f_1, f_2, f_3\}$ is a basis for V^*, and then find a basis for V for which it is the dual basis.

4. **Let n be a positive integer and F a field. Let W be the set of all vectors $(x_1, x_2, \ldots, x_n)$ in F^n such that $x_1 + x_2 + \cdots + x_n = 0$.**
   (a) **Prove that W^0 consists of all linear functionals f of the form $f(x_1, x_2, \ldots, x_n) = c\sum_{j=1}^{n} x_j$,**
   (b) **Show that the dual space W^* of W can be naturally identified with the linear functionals $f(x_1, x_2, \ldots, x_n) = c_1x_1 + c_2x_2 + \cdots + c_nx_n$ on F^n which satisfy $c_1 + c_2 + \cdots + c_n = 0$**

5. **Let S be a set, F a field, and $V(S; F)$ the space of all functions from S into F;**
   $(f + g)(x) = f(x) + g(x)$
   $(cf)(x) = cf(x)$
   Let W be any n-dimensional subspace of $V(S; F)$. Show that there exist points $x_1, x_2, \ldots, x_n$ in S and functions $f_1, f_2, \ldots, f_n$ in W such that
   $f_i(x_j) = \begin{cases} 0\ for\ i \neq j \\ 1\ for\ i = j \end{cases}$.

2.13 EIGENSPACE OF A LINEAR OPERATOR

Let us recall the concept of eigenvalue and eigenvector of a linear operator.

Let T be a linear operator on a vector space V. A nonzero vector $v \in V$ is called an **eigenvector** of T if there exists a scalar λ such that $T(v) = \lambda v$. The scalar λ is called the **eigenvalue** corresponding to the eigenvector v.

Let T be a linear operator on an n-dimensional vector space V with an ordered basis B. We define the **characteristic polynomial** $f(t)$ of T to be the characteristic polynomial of A, the matrix representation of T with respect to the ordered basis B, that is,

$$f(t) = \det(A - tI_n).$$

We often denote the characteristic polynomial of an operator T by $\det(T - tI)$.

The equation $f(t) = 0$ is known as **characteristic equation** of T or of the matrix A as defined above. The roots of the characteristic equation $f(t) = 0$ are known as eigenvalues of T. If λ be a root of $f(t) = 0$, that is, λ is an eigenvalue of

a linear operator T, then the **algebraic multiplicity** of λ is the largest posi integer k for which $(t-\lambda)^k$ is a factor of $f(t)$.

2.13.1 Definition. Let T be a linear operator on a vector space V and λ be a eigenvalue of T. Define

$$E_\lambda = \{ v \in V : T(v) = \lambda v \}$$

The set E_λ is called the **eigenspace**of T corresponding to the eigenvalue λ.

It is to be noted that E_λ is the null space of $T-\lambda I$. Clearly, E_λ is the collection of all eigenvectors of T corresponding to λ together with null vector. It is easy to see that E_λ is a subspace of V. The number of elements in a basis of E_λ, that is, the maximum number of linearly independent eigenvectors of T corresponding to the eigenvalue λ is the dimension of E_λ.

Dimension of E_λ is called the **geometric multiplicity of** λ.

Now, we shall prove that the geometric multiplicity of λ is always less equal to algebraic multiplicity of λ.

2.13.2Theorem. Let T be a linear operator on a finite dimensional vector space V and let λ be an eigenvalue of T having algebraic multiplicity m. Then

$$1 \le \dim E_\lambda \le m, \text{ that is,}$$

$$1 \le \textit{geometric multiplicity of } \lambda \le \textit{algebraic multiplicity of } \lambda$$

Proof. Let us consider an ordered basis $\{v_1, v_2, \dots, v_k\}$ for E_λ. Since, $\{v_1, v_2, \dots, v_k\}$ is an independent set in $E_\lambda (\subseteq V)$, it can be extended to an ordered basis$B = \{v_1, v_2, \dots, v_k, \dots, v_n\}$ for V. Let A be the matrix representation of T with respect to the ordered basis B. Thus, each $v_i (1 \le i \le k)$ is an eigenvector of T corresponding to λ and hence,

$$A = \begin{pmatrix} \lambda I_k & B \\ 0 & C \end{pmatrix}$$

where B and C are square matrices.

The characteristic polynomial of T is given by

$$f(t) = \det(A - tI_n) = \det\begin{pmatrix} (\lambda - t)I_k & B \\ 0 & C - tI_{n-k} \end{pmatrix}$$

$$= \det\big((\lambda - t)I_k\big)\det(C - tI_{n-k})$$

$$= (\lambda - t)^k g(t)$$

where $g(t)$ is a polynomial. Thus, $(\lambda - t)^k$ is a factor of $f(t)$, and hence the algebraic multiplicity of λ is at least k, i.e., $m \ge k$. But $\dim E_\lambda = k$. Hence,

$$1 \le k \le m \blacksquare$$

2.13.3 Example.

Let T be the linear operator on $\mathbb{R}^3$ defined by

$$T(x, y, z) = (4x + z, 2x + 3y + 2z, x + 4z)$$

Let us consider the ordered basis $\{(1,0,0), (0,1,0), (0,0,1)\}$ for $\mathbb{R}^3$. Now,

$$T(1,0,0) = (4,2,1), \qquad T(0,1,0) = (0,3,0) \qquad T(0,0,1) = (1,2,4)$$

Hence, the matrix representation A of T with respect to this ordered basis is given by

$$A = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$$

and hence the characteristic polynomial of T, that is, of A is

$$f(t) = \det(A - tI) = \det\begin{pmatrix} 4-t & 0 & 1 \\ 2 & 3-t & 2 \\ 1 & 0 & 4-t \end{pmatrix} = -(t-5)(t-3)^2 .$$

So, the eigenvalues of T are $\lambda_1 = 5$ and $\lambda = 3$ with algebraic multiplicities 1 and 2 respectively. Now, if E_{λ_i} be the eigenspace of T corresponding to λ_i $(i = 1,2)$ then

$$E_{\lambda_1} = N_{T-\lambda_1 I} = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} : \begin{pmatrix} 4-5 & 0 & 1 \\ 2 & 3-5 & 2 \\ 1 & 0 & 4-5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\}$$

$$= \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} : \begin{pmatrix} -1 & 0 & 1 \\ 2 & -2 & 2 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

Thus, we get a system of linear equations

$$\begin{aligned} -x \qquad\quad + z &= 0 \\ 2x - 2y + 2z &= 0 \\ x \qquad\quad - z &= 0 \end{aligned}$$

Solving, we get $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = c \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$, c is a scalar.

Hence, $E_{\lambda_1} = \left\{ c \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} : c \in F \right\}$.

Since, any single non-null vector is linearly independent, we see that $\left\{\begin{pmatrix}1\\2\\1\end{pmatrix}\right\}$ is a basis for E_{λ_1}.

Hence, $\dim E_{\lambda_1} = 1$, in other words, geometric multiplicity of λ_1 is 1.

Similarly, $E_{\lambda_2} = N_{T-\lambda_2 I}$ is given by

$$E_{\lambda_2} = \left\{\begin{pmatrix}x\\y\\z\end{pmatrix} : \begin{pmatrix}4-3 & 0 & 1\\ 2 & 3-3 & 2\\ 1 & 0 & 4-3\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}\right\}$$

That is, the solution space of the system

$$x + z = 0$$
$$2x + 2z = 0$$
$$x + z = 0$$

Solving, $x = -z$. Taking $z = t$, we have, $x = -t$. Now, there is no y in the system, so y can take any arbitrary value. Let us take $y = s$. Thus, we have,

$$\begin{pmatrix}x\\y\\z\end{pmatrix} = s\begin{pmatrix}0\\1\\0\end{pmatrix} + t\begin{pmatrix}-1\\0\\1\end{pmatrix},\quad s,t \in \mathbb{R}.$$

Therefore,

$$E_{\lambda_2} = \left\{s\begin{pmatrix}0\\1\\0\end{pmatrix} + t\begin{pmatrix}-1\\0\\1\end{pmatrix} : s,t \in \mathbb{R}\right\} = L(W)$$

where $W = \left\{\begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}-1\\0\\1\end{pmatrix}\right\}$. Since, W is linearly independent, W is a basis of E_{λ_2}.

Hence, $\dim E_{\lambda_2} = 2$. In other words, geometric multiplicity of λ_2 is 2.

2.14 DIAGONALIZABILITY

Recall that an $n \times n$ matrix A is similar to an $n \times n$ matrix B if there exists a non-singular matrix P such that $B = P^{-1}AP$ or $B = PAP^{-1}$.

2.14.1 Definition. A square matrix A of order n is **diagonalizable** if it is similar to a diagonal matrix of order n.

Thus $A = (a_{ij})_{n\times n}$ is diagonalizable if there exists a non-singular matrix P such that

$$P^{-1}AP = D = diag\,\{\lambda_1, \lambda_2, \ldots, \lambda_n\}.$$

$D = diag\,\{\lambda_1, \lambda_2, \ldots, \lambda_n\} = (d_{ij})_{n\times n}$ is a matrix where $d_{ij} = 0$ for $i \neq j$ and $d_{ii} = \lambda_i$ for $i = 1, 2, \ldots, n$.

But one question may be raised, my dear readers. *Why should we try to diagonalize a given $n \times n$ matrix?* Let us try to find the answer. Suppose, A is a diagonalizable matrix (not all square matrices are diagonalizable, we will come to that point later), then it is easy to find A^k for some positive integer k.

Let's try to explain. If A is diagonalizable then there exists a non-singular matrix P such that

$$P^{-1}AP = D = diag\,\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$$

What is $D^k, k \in \mathbb{Z}^+$? We have, $D^k = diag\,\{\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k\}$. again, we see that,

$$A = PDP^{-1}.$$

Then, $A^2 = (PDP^{-1})(PDP^{-1}) = PD^2P^{-1}$.

Similarly, $A^k = (PDP^{-1})(PDP^{-1}) \ldots (PDP^{-1})[k\ times] = PD^kP^{-1}$

Since each P^{-1} cancels an P, except for the first P and the last P^{-1}. Thus,

$$A^k = PD^kP^{-1} = P.diag\,\{\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k\}.P^{-1}$$

which is easy to compute.

Example : Let $A = \begin{pmatrix}1 & 5\\0 & 6\end{pmatrix}$. Let $P = \begin{pmatrix}1 & 1\\0 & 1\end{pmatrix}$. (How this P comes? Will be discussed later).

Eigenvalues of A are 1,6. Hence, we have,

$$A^k = \begin{pmatrix}1 & 5\\0 & 6\end{pmatrix}^k = \begin{pmatrix}1 & 1\\0 & 1\end{pmatrix}\begin{pmatrix}1^k & 0\\0 & 6^k\end{pmatrix}\begin{pmatrix}1 & -1\\0 & 1\end{pmatrix} = \begin{pmatrix}1 & 6^k - 1\\0 & 6^k\end{pmatrix}.$$

Thus, it is easy to find A^5 or A^{50} or any positive integral power of A. Since, $\det A \neq 0$, A^{-1} exists. Therefore, putting $k = -1$, we have,

$$A^{-1} = \begin{pmatrix}1 & 6^{-1} - 1\\0 & 6^{-1}\end{pmatrix} = \begin{pmatrix}1 & -\frac{5}{6}\\0 & \frac{1}{6}\end{pmatrix}.$$

Now, we are ready to face big questions? *Which matrices are diagonalizable and how can we diagonalize a given diagonalizable matrix ?* But before that we need some preparedness.

2.14.2 Theorem. If $A \in M_{n\times n}(F)$ is diagonalizable, then the characteristic polynomial $f(t)$ of A is of the form $f(t) = (t-\lambda_1)(t-\lambda_2)\dots(t-\lambda_n)$ for some $\lambda_i \in F, i = 1,2,..,n$.

Proof. Since A is diagonalizable, it is similar to a diagonal matrix D where

$$P^{-1}AP = D = diag\{\lambda_1, \lambda_2, \dots, \lambda_n\}, \qquad \lambda_i \in F, i = 1,2,..,n.$$

for some non-singular matrix P.

Then $tI_n - D = diag\{t-\lambda_1, t-\lambda_2, \dots, t-\lambda_n\}$.

Thus, $\det(tI_n - D) = (t-\lambda_1)(t-\lambda_2)\dots(t-\lambda_n)$.

Now,

$$\det(tI_n - D) = \det(tP^{-1}P - P^{-1}AP) = \det(P^{-1}(tI_n - A)P)$$
$$= \det P^{-1}\det(tI_n - A)\det P = \det(tI_n - A)$$

Hence, $\det(tI_n - A) = (t-\lambda_1)(t-\lambda_2)\dots(t-\lambda_n)$.

Note. Remember that λ_i's in the theorem 2.14.2 may not be distinct, that is, the characteristic polynomial of a diagonalizable matrix A could be in the form

$$\det(tI_n - A) = (t-\lambda_1)^{d_1}(t-\lambda_2)^{d_2}\dots(t-\lambda_k)^{d_k}$$

This theorem also shows that if the characteristic polynomial of a given square matrix cannot be expressed as in the form given above, then the matrix is *not diagonalizable*. For example, let

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_{n\times n}(\mathbb{R})$$

Then the characteristic polynomial $f(t)$ of A is given by

$$f(t) = \det(tI_n - A) = \det\begin{pmatrix} t-0 & -1 \\ 1 & t-0 \end{pmatrix} = t^2 + 1.$$

Thus, $f(t)$ cannot be expressed as a product of linear factors over $\mathbb{R}$. Hence, by theorem 2.14.2 the given matrix A is not diagonalizable.

But one more thing is to be mentioned here that if the characteristic polynomial of a given square matrix have linear factors, it is not necessary that A is diagonalizable, that is, converse of the theorem 2.14.2 is not true. For example, let

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Then the characteristic polynomial $f(t)$ of A is given by

$$f(t) = (t-1)^2.$$

Now, if A is diagonalizable then A is similar to some diagonal matrix $D = \begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix}$.

Thus,

$$(t-1)(t-1) = \det(tI_2 - D) = \det\begin{pmatrix} t-c & 0 \\ 0 & t-d \end{pmatrix} = (t-c)(t-d).$$

Therefore, we have, $c = 1 = d$. Thus, $D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2$.

Now, there exists a non-singular matrix P such that $P^{-1}AP = D = I_2$.

Therefore,

$$A = PI_2P^{-1} = I_2$$

which is a contradiction. Hence, A cannot be similar to a diagonal matrix. In other words, A is not diagonalizable.

Now, let us try to find when a square matrix is diagonalizable.

2.14.3 Theorem. An $n \times n$ matrix A over a field F is diagonalizable if and only if there exists n eigenvectors of A which are linearly independent.

Proof. Let A be an $n \times n$ matrix over a field F.

Let A be diagonalizable, that is, A is similar to a diagonal matrix $D = diag\{d_1, d_2, \dots, d_n\}$.

So, there exists a non-singular matrix $P = (p_{ij})_{n\times n}$ such that $P^{-1}AP = D$, that is, $AP = PD$.

Then jth column vector of AP = the jth column vector of PD. Now, the jth column vector of AP is

$$A\begin{pmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{pmatrix}$$

and that of PD is $d_j\begin{pmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{pmatrix}$. Hence,

$$A\begin{pmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{pmatrix} = d_j\begin{pmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{pmatrix}.$$

Thus, we see that the jth column vector of P is an eigenvector of A corresponding to the eigenvalue d_j of A.

Thus, each column vector of P is an eigenvector of A. Since, P is non-singular, it is clear that all these n eigenvectors, that is, the column vectors of A are linearly independent.

Conversely, let $v_1, v_2, \dots, v_n$ be n linearly independent eigenvectors corresponding to the respective eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, all of which may not be distinct. Here

$$v_j = \begin{pmatrix} v_{1j} \\ v_{2j} \\ \vdots \\ v_{nj} \end{pmatrix}.$$

Let P be the $n \times n$ matrix whose jth column vector is v_j. That is,

$$P = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1j} & \dots & v_{1n} \\ v_{21} & v_{22} & \dots & v_{2j} & \dots & v_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ v_{n1} & v_{n2} & \dots & v_{nj} & \dots & v_{nn} \end{pmatrix}$$

Since, $v_1, v_2, \dots, v_n$ are linearly independent, we see that P is non-singular.

Let $D = diag\,\{\lambda_1, \lambda_2, \dots, \lambda_n\}$.

Now, the jth column vector of AP is Av_j and that of PD is $\lambda_j v_j$.

Since, $Av_j = \lambda_j v_j$, we see that the jth column vector of AP equals the jth column vector of PD for $j = 1,2, \dots n$. Hence, $AP = PD$. Since, P is non-singular, we have, $P^{-1}AP = D$.

Thus, A is similar to a diagonal matrix D. Therefore, A is diagonalizable. Hence the theorem.

Corollary : If eigenvalues of a square matrix $A \in M_{n\times n}(F)$ are all distinct then A is diagonalizable.

Proof. We first show that eigenvectors corresponding to distinct eigenvalues of a square matrix A are linearly independent.

Let X_1 and X_2 be two eigenvectors corresponding to two distinct eigenvalues λ_1 and λ_2 respectively. Then, we have, $AX_1 = \lambda_1 X_1$ and $AX_2 = \lambda_2 X_2$. $\lambda_1 \neq \lambda_2$, $X_1 \neq 0$, $X_2 \neq 0$

Let $c_1X_1 + c_2X_2 = 0 \dots (1)$ where $c_1, c_2 \in F$. Then

$$c_1AX_1 + c_2AX_2 = 0 \;\; i.e. \;\; c_1\lambda_1X_1 + c_2\lambda_2X_2 = 0 \dots (2)$$

If we multiply the equation (1) by λ_2 and subtract from the equation (2), we get, $c_1(\lambda_1 - \lambda_2)X_1 = 0$ which shows that $c_1 = 0$ as $\lambda_1 \neq \lambda_2$ and $X_1 \neq 0$.

If $c_1 = 0$, then from equation (1), we have $c_2 = 0 (X_2 \neq 0)$, that is, X_1, X_2 are linearly independent.

Hence, by the theorem, A is diagonalizable ■

Remember that *there is no connection between invertibility and diagonalizability.* A square matrix A is invertible if and only if all its eigenvalues are non-zero. Thus invertibility is concerned with eigenvalues of the matrix. But A is diagonalizable if and only if A has enough (equating the order of the matrix) number of eigenvectors. In other words, if the number of independent eigenvectors is less than the order of the matrix then the matrix is not diagonalizable. Therefore, diagonalizibility of a square matrix A is concerned with the number of independent eigenvectors of A.

2.15 INVARIANT SUBSPACES

Any linear mapping from a finite dimensional vector space V to itself is called a **linear operator on V**. Let $L(V)$ be the collection of all linear operators on V. Now, time has come to introduce an idea of *invariant subspace.* If T is a linear operator on a finite dimensional vector space V ($i.e. T : V \to V$ *is linear*) then a subspace U of V is called an **invariant subspace under T** if $u \in U \Rightarrow Tu \in U$. Clearly, the zero subspace $\{0_V\}$ and the vector space itself are invariant subspaces under any linear operator T. It is also clear that $kerT$ and ImT are invariant as

$$u \in kerT \Rightarrow Tu = 0 \Rightarrow T(Tu) = T(0) = 0 \Rightarrow Tu \in kerT$$

and $$v \in ImT \Rightarrow Tv \in ImT.$$

Although $kerT$ and ImT are invariant subspaces under T of V but $kerT$ could be $\{0_V\}$ or ImT could be V itself. We will try to face the problem 'Is there any non-trivial invariant subspace of dimension 1?' How does an operator behave on an invariant subspace of dimension 1? Subspaces of V of dimension 1 are easy to describe. Take any nonzero vector $u \in V$ and let U equal the set of all scalar multiples of u ,that is,

$$U = \{cu : c \in F\}$$

Then U isaone-dimensionalsubspaceof V, andeveryone-dimensional subspace of V is of this form. If $u \in V$ and the subspace$U$, given above, is invariant under $T \in L(V)$, then Tu must be in U, and hence there must be a scalar $\lambda \in F$ such that $Tu = \lambda u$. Conversely, if u is a nonzero vector in V such that $Tu = \lambda u$ for some $\lambda \in F$, then the subspace U defined above, isaone-dimensionalsubspaceof V invariant under T.

The equation $Tu = \lambda u$

which we have just seen is intimately connected with one-dimensional invariantsubspaces,isimportantenoughthatthevectors$u$andscalars λ satisfying it

are given special names. Specifically, a scalar $\lambda \in F$ is called an **eigenvalue of** $T \in L(V)$ if there exists a nonzero vector u such that $Tu = \lambda u$. We must require u to be nonzero because with $u = 0$ every scalar $\lambda \in F$ satisfies the above equation. Thus we see that T has a one-dimensional invariant subspace if and only if T has an eigenvalue.

Look, the equation $Tu = \lambda u$ is equivalent to $(T - \lambda I)u = 0$ and we know that $(T - \lambda I)(0) = 0$, thus it is clear that $T - \lambda I$ is not injective, that is, the operator $T - \lambda I$ is not invertible (singular).

Suppose $T \in L(V)$ and $\lambda \in F$ is an eigenvalue of T. A nonzero vector $u \in V$ is called an **eigenvector** of T (corresponding to λ) if $Tu = \lambda u$. Since it is equivalent to $(T - \lambda I)u = 0$, we see that the set of eigenvectors of T corresponding to λ together with zero vector equals $ker(T - \lambda I)$. In particular, the set of eigenvectors of T corresponding to λ, together with null vector, is a subspace of V. Since for an eigenvalue λ there exists a nonzero vector u such that $Tu = \lambda u$, that is, $(T - \lambda I)u = 0$, we can say that $T - \lambda I$ is not injective as we know, $(T - \lambda I)0 = 0$ but $u \neq 0$. Thus, λ is an eigenvalue of T if and only if $T - \lambda I$ is singular, that is, if and only if $T - \lambda I$ is not invertible and this happens if and only if $T - \lambda I$ is not injective.

The main reason that theories for operators richer than the theories of linear maps is that operators can be raised to powers.

Since an operator T on V is a linear map from V to V, for any $v \in V$, we have, $T(v) \in V$. Then we get $T(T(v)) \in V$, $T\left(T(T(v))\right) \in V$ and so on. We prefer to write T^2 instead of TT and in general, for a positive integer m, we can define T^m by

$$T^m = TT \ldots.T \ (m \ times).$$

For convenience, we define, T^0 to be the identity operator I on V.

If T is an invertible operator, then inverse of T is denoted by T^{-1}. For a positive integer m, we define T^{-m} to be $(T^{-1})^m$.

One can easily verify that if T is an operator, then

$$T^m T^n = T^{m+n} \ and \ (T^m)^n = T^{mn}$$

where m, n are integers if T is invertible and m, n are nonnegative integers if T is not invertible.

Now, if $p \in \mathcal{P}(F)$ is a polynomial given by

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for $z \in F$.

If $T \in L(V)$ then the operator $p(T)$ is defined

$$p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_m T^m.$$

For example, if p is the polynomial defined by $p(z) = 2z^3 + 5z + 4$ for $z \in F$, then

$$p(T) = 2T^3 + 5T + 4I.$$

Look ! This is a new use of the symbol p because we are applying it to operators, not just element of F. It is easy to see that for some fixed operator $T \in L(V)$, the function from $\mathcal{P}(F)$ to $L(V)$ given by $p \to p(T)$ is linear.

If p and q are polynomials with coefficients in F, that is, if $p, q \in \mathcal{P}(F)$, then pq is the polynomial defined by

$$(pq)(z) = p(z)q(z)$$

for $z \in F$. Similarly, we have,

$$(pq)(T) = p(T)q(T)$$

for all $T \in L(V)$.

2.16.1 Cayley- Hamilton Theorem

Let A be an $n \times n$ square matrix. Let

$f(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_0$ be the characteristic polynomial of A.

Then $c_n A^n + c_{n-1} A^{n-1} + \cdots + c_0 I_n = 0$

Thus, A satisfies its characteristic equation.

Proof. We know that for any square matrix P,

$$P \ adjP = \det P \ I.$$

Applying this result to the matrix $A - xI$, we get

$$(A - xI)adj\ (A - xI) = \det(A - xI)I = \chi_A(x)I \ \ldots\ldots (1)$$

Now, $adj\,(A - xI)$ is a matrix whose entries are determinants (up to sign) of $(n-1)$ square submatrices of $A - xI$. Hence $adj\ (A - xI)$ is a matrix whose entries are polynomials in x of degree at most $n - 1$. Thus,

$$adj\ (A - xI) = B_{n-1}x^{n-1} + B_{n-2}x^{n-2} + \cdots + B_1 x + B_0$$

where $B_i (i = 0,1, \ldots, n-1)$ are matrices with real entries. Hence, equation (1) can be written as

$$\begin{aligned}(A - xI)(B_{n-1}x^{n-1} + B_{n-2}x^{n-2} + \cdots + B_1 x + B_0) \\ = (a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0)I \\ = (a_n I)x^n + (a_{n-1}I)x^{n-1} + \cdots + (a_1 I)x + a_0 I\end{aligned}$$

Comparing the coefficients of like powers of x, we get

$$-B_{n-1} = a_n I$$
$$AB_{n-1} - B_{n-2} = a_{n-1} I$$
$$AB_{n-2} - B_{n-3} = a_{n-2} I$$
$$\dots\dots\dots\dots$$
$$AB_0 = a_0 I$$

Multiplying the first of these equations by A^n, the second by A^{n-1}, so on, the last but one by A and the last one by I, and adding them we get

$$a_n A^n + a_{n-1}A^{n-1} + \dots + a_1 A + a_0 I = 0$$

which completes the proof.

2.16.2 Cayley-Hamilton Theorem for Linear Operator

Let T be a linear operator on a finite dimensional vector space V over a field F with $f(x)$ as its characteristic polynomial. Then $f(T) = 0$, 0 being the zero transformation. That is, T satisfies its characteristic equation.

Proof. Let B be an ordered basis for V. Let A be the matrix representation of T with respect to the ordered basis B. Then

$$f(x) = \det(A - XI_n).$$

By theorem 2.16.1, $$f(A) = 0.$$

By the isomorphism between $\mathcal{L}(V)$ and $M_{n\times n}(F)$, we get $f(T) = 0$.

Solved examples :

1. **Let V be an n −dimensional vector space over a field F. What is the characteristic polynomial of the identity operator on ? What is the characteristic polynomial for the zero operator ?**

Solution. The identity operator on V can be represented by the $n\times n$ identity matrixI_n. Thus the characteristic polynomial of the identity operator is given by

$$\det(xI_n - I_n) = (x-1)^n.$$

The zero operator on V can be represented by $n\times n$ zero matrix 0_n. Thus its characteristic polynomial is given by

$$\det(xI_n - 0_n) = x^n.$$

2. **Prove that 1 is an eigenvalue of every square matrix with the property that the sum of the entries in each row equals 1.**

Solution. Let A be an $n\times n$ matrix such that the sum of the entries in each row is 1.

Thus, $A = (a_{ij})_{n\times n}$ where $\sum_{j=1}^{n} a_{ij} = 1$, for $i = 1,2,..,n$. Let

$$v = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}_{n\times 1}$$

Then by matrix multiplication rule, we have,

$$Av = \begin{pmatrix} \sum_{j=1}^{n} a_{1j} \\ \sum_{j=1}^{n} a_{2j} \\ \dots \\ \sum_{j=1}^{n} a_{nj} \end{pmatrix}_{n\times 1} = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}_{n\times 1} = v.$$

Thus, $Av = v$ shows that 1 is an eigenvalue of A.

3. **Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Suppose $a, b \in \mathbb{R}$ such that$T^2 + aT + bI = 0$. Prove that T has a real eigenvalue if and only if$a^2 \geq 4b$.**

Solution. Let us first suppose that T has an eigenvalue $\lambda \in \mathbb{R}$. Then there exists a non-zero vector v such that$Tv = \lambda v$.

Thus, $T^2 v = T(Tv) = T(\lambda v) = \lambda(Tv) = \lambda(\lambda v) = \lambda^2 v$. Then

$(T^2 + aT + bI)v = 0 \Rightarrow \lambda^2 v + \lambda a v + bv = 0 \Rightarrow (\lambda^2 + a\lambda + b)v = 0$

Since, $v \neq 0$, we have, $\lambda^2 + a\lambda + b = 0$. For real value of λ, we have,

$$a^2 - 4b \geq 0 \;\; i.e. \;\; a^2 \geq 4b.$$

Conversely, let $a^2 \geq 4b$. Then the equation $\lambda^2 + a\lambda + b = 0$ has real roots say λ_1, λ_2. Then

$$\lambda^2 + a\lambda + b = (\lambda - \lambda_1)(\lambda - \lambda_2).$$

Hence,$0 = T^2 + aT + bI = (T - \lambda_1 I)(T - \lambda_2 I)$ which shows that$(T - \lambda_1 I)(T - \lambda_2 I)$ is not injective. In other words, at least one of $T - \lambda_1 I$ and$T - \lambda_2 I$ is not injective. So, at least one of λ_1 and λ_2 must be an eigenvalue of T. Hence, T has a real eigenvalue.

4. **Define $T \in \mathcal{L}(\mathbb{C}^2)$ by $T(w,z) = (z,0)$. Find the set of eigenvectors of T.**

Solution. Let λ be an eigenvalue of T. Then the eigenvalue-eigenvector equation is given by

$$T(w,z) = \lambda(w,z)$$

i.e. $$(z,0) = \lambda(w,z).$$

Thus, we have,$z = \lambda w$ and $0 = \lambda z$.

If $\lambda \neq 0$, then $z = 0$ and hence $w = 0$, that is, $(w, z) = (0,0)$.

Since an eigenvalue must have a non-zero eigenvector, we may conclude that λ must be zero, that is, 0 is the only possible eigenvalue of T. Now, if $\lambda = 0$ then z must be zero but w can take any arbitrary value. Thus, 0 is the only eigenvalue of T and the set of eigenvectors corresponding to 0 is given by

$$\{(w, 0) : w \in \mathbb{C} - \{0\}\}$$

5. Suppose $N \in \mathcal{L}(V)$ is nilpotent. Prove that 0 is the only eigenvalue of T.

Solution. Let m be a positive integer such that $N^m = 0$, that is, $N^m(v) = 0$, $\forall v \in V$. Thus, N is not injective. So, 0 is an eigenvalue of N. We shall show that 0 is the only eigenvalue of N.

Let λ be an eigenvalue of N. Then there exists a non-zero vector $v \in V$ such that $N(v) = \lambda v$

Therefore, $N^2(v) = N(N(v)) = N(\lambda v) = \lambda N(v) = \lambda^2 v$.

After repeated application of N on both sides we have,

$$N^m(v) = \lambda^m v$$

$$\Rightarrow 0 = \lambda^m v \Rightarrow \lambda = 0 \ (as\ v \neq 0).$$

Hence, 0 is the only eigenvalue of N.

6. Let T be a linear operator on $\mathbb{R}^3$ which is represented in the standard ordered basis by the matrix

$$\begin{pmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{pmatrix}.$$

Prove that T is diagonalizable by exhibiting a basis for $\mathbb{R}^3$, each vector of which is a characteristic vector of T.

Solution. Let

$$A = \begin{pmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{pmatrix}$$

Thus the characteristic polynomial of A is given by

$$\det(xI_n - A)$$

$$= det \begin{pmatrix} x+9 & -4 & -4 \\ 8 & x-3 & -4 \\ 16 & -8 & x-7 \end{pmatrix}$$

$$= (x+1)^2(x-3) \quad [simplify]$$

Thus the eigenvalues of A are given by $\lambda_1 = -1, \ \lambda_2 = 3$.

For $\lambda_1 = -1$, we have,

$$(\lambda_1 I_n - A)X = 0$$

$$\Rightarrow \begin{pmatrix} 8 & -4 & -4 \\ 8 & -4 & -4 \\ 16 & -8 & -8 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

i.e. $8x - 4y - 4z = 0$ and $16x - 8y - 8z = 0$ which gives $2x - y - z = 0$. Hence,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 2x - y \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \qquad x, y \in \mathbb{R}.$$

Thus, corresponding to the eigenvalue -1, we get two eigenvectors $\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$.

Again, for the eigenvalue $\lambda_2 = 3$, the equation $(\lambda_2 I - A)X = 0$ gives

$$\begin{pmatrix} 12 & -4 & -4 \\ 8 & 0 & -4 \\ 16 & -8 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

This gives $12x - 4y - 4z = 0, \ 8x - 4z = 0, \ 16x - 8y - 4z = 0$ which gives

$$z = 2x, \qquad y = x$$

Hence, $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ x \\ 2x \end{pmatrix} = x \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \ x \in \mathbb{R}.$

Hence, an eigenvector corresponding to the eigenvalue 3 is given by $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$.

Since

$$\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 2 & -1 & 2 \end{vmatrix} \neq 0$$

we see that the vectors $\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ are linearly independent.

Hence, A is diagonalizable. In other words. T is diagonalizable.

7. Let A and B be $n \times n$ matrices over the field F. Prove that if $I - AB$ is invertible, then $I - BA$ is invertible and

$$(I - BA)^{-1} = I + B(I - AB)^{-1}A.$$

Solution. We have,

$$(I - BA)[I + B(I - AB)^{-1}A]$$
$$= I + B(I - AB)^{-1}A - BA - BAB(I - AB)^{-1}A$$
$$= I + B[(I - AB)^{-1} - I - AB(I - AB)^{-1}]A$$
$$= I + B\,[(I - AB)^{-1}(I - AB) - I]A$$
$$= I + B(I - I)A = I$$

Hence, $(I - BA)^{-1} = I + B(I - AB)^{-1}A.$

Exercise

1. Consider a 2×2 matrix of real numbers

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Prove that A has an eigenvalue in $\mathbb{R}$ if and only if $(a - d)^2 + 4bc \geq 0$.

2. Let

$$A = \begin{pmatrix} 6 & -3 & -2 \\ 4 & -1 & -2 \\ 10 & -5 & -3 \end{pmatrix}$$

Is A diagonalizable over the field $\mathbb{R}$?

3. Let T be the linear operator on $\mathbb{R}^4$ which is represented in the standard ordered basis by the matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \end{pmatrix}.$$

4. Let T be a linear operator on the n-dimensional vector space V, and suppose that T has n distinct eigenvalues. Prove that T is diagonalizable.

2.17 THE MINIMAL POLYNOMIAL OF A LINEAR OPERATOR

First of all let us try to understand monic polynomial.

2.17.1 Definition. A monic polynomial is a polynomial whose highest degree coefficient is 1.

For example, $z^3 + 2z + 3$ is a monic polynomial but $2z^3 + z + 1$ is not.

Let V be a vector space over a field F with $\dim V = n$. Let $\mathcal{L}(V)$ be the space containing all linear operators on V. Therefore, $\dim \mathcal{L}(V) = n^2$. Then the set

$$S = \{I, T, T^2, \ldots, T^{n^2}\}$$

cannot be linearly independent as S contains $n^2 + 1$ elements (remember, a linearly independent set in a vector space with dimension n^2 can contain atmost n^2 elements).

Let m be the smallest positive integer such that

$$\{I, T, T^2, \ldots, T^m\}$$

is linearly dependent. Since this set is linearly dependent, we know that one of the operators in the list above is a linear combination of the previous ones. Because m was chosen to be smallest positive integer such that the set given above is linearly dependent, we conclude that T^m is a linear combination of $I, T, T^2, \ldots, T^{m-1}$. So, there exist scalars $a_0, a_1, \ldots, a_{m-1} \in F$ such that

$$a_0 I + a_1 T + a_2 T^2 + \cdots a_{m-1} T^{m-1} + T^m = 0.$$

The choice of scalars $a_0, a_1, \ldots, a_{m-1} \in F$ above is unique because if there are two choices like

$$a_0 I + a_1 T + a_2 T^2 + \cdots a_{m-1} T^{m-1} + T^m = 0$$

and $b_0 I + b_1 T + b_2 T^2 + \cdots b_{m-1} T^{m-1} + T^m = 0$

for $b_0, b_1, \ldots, b_{m-1} \in F$, then after subtraction we get

$$(a_0 - b_0)I + (a_1 - b_1)T + \cdots + (a_{m-1} - b_{m-1})T^{m-1} = 0$$

Linear independence of $\{I, T, T^2, \ldots, T^{m-1}\}$ shows that $a_i = b_i$ for $i = 0, 1, \ldots, m - 1$.

The polynomial $p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$ is called the **minimal polynomial** of T.

It is the monic polynomial $p \in \mathcal{P}(F)$ of smallest degree such that $p(T) = 0$.

For a matrix $A \in M_{n \times n}(F)$, the **minimal polynomial** $p(z)$ of A is the monic polynomial of least positive degree for which $p(A) = 0$, 0 being the null matrix.

For example, the minimal polynomial of the identity operator I is $z - 1$ and the minimal polynomial of the operator on F^2 whose matrix equals to $\begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}$ is

$$p(z) = 6 - 5z + z^2$$

Now, let us try to remember division algorithm for polynomials, which states that

Suppose $f, g \in \mathcal{P}(F)$, that is, f and g be two polynomials, with $g \neq 0$. Then there exist polynomials $q, r \in \mathcal{P}(F)$ such that

$f = gq + r$ *where* $0 \leq \deg r < \deg g$.

If $r = 0$ i.e. if $f = gq$ then we say that the polynomial g divides the polynomial f.

By Cayley-Hamilton theorem for linear operators, we know that, for a linear operator T defined on an n −dimensional vector space V, there is a polynomial $f(x)$ of degree n , known as characteristic polynomial of T, such that $f(T) = 0$, 0 being the zero transformation. Therefore, the degree of the minimal polynomial of each operator on V has degree at most n, that is, $\dim V$. Why? The next theorem will clarify it.

2.17.2 Theorem. Let $p(z)$ be the minimal polynomial of a linear operator T on a finite-dimensional vector space V. Then for any polynomial $g(z)$, if $g(T) = 0$, 0 being the zero transformation, then $p(z)$ divides $g(z)$. In particular, $p(z)$ divides the characteristic polynomial of T.

Proof. Let $g(z)$ be a polynomial such that $g(T) = 0$, 0 being zero transformation. Then, by the division algorithm, there exist polynomials $q(z)$ and $r(z)$ such that

$$g(z) = q(z)p(z) + r(z)$$

where $0 \leq \deg r < \deg p$. Since, $g(T) = 0$, we have,

$$q(T)p(T) + r(T) = 0.$$

Again, $p(T) = 0$ gives $r(T) = 0$. Since, degree of $r(z)$ is less than the degree of $p(z)$ and $p(z)$ is the minimal polynomial of T, $r(z)$ must be the zero polynomial.

Thus, we have, $g(z) = q(z)p(z)$ which shows that $p(z)$ divides $g(z)$.

In particular, if we take $g(z)$ as the characteristic polynomial of T, then the last part is proved■

Now we describe the eigenvalues of an operator in terms of its minimal polynomial.

2.17.3 Theorem. Let T be a linear operator defined on finite dimensional vector space V. Then the zeros of the minimal polynomial of T are precisely the eigenvalues of T.

Proof. Let $p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$ be the minimal polynomial of T.

First suppose that $\lambda \in F$ is a zero of p. Then, we have,

$$p(z) = (z - \lambda)q(z)$$

where q is a monic polynomial with coefficients in F. Now, $p(T) = 0$ gives

$$0 = (T - \lambda I)q(T)$$

i.e. $0 = (T - \lambda I)q(T)(v)$, $\forall v \in V$.

Since, the degree of q is less than the degree of the minimal polynomial p, $q(T) \neq 0$, 0 being the zero transformation. In other words, there exists $u \in V$ such that $q(T)(u) \neq 0$.

Hence, $(T - \lambda I)q(t)(u) = 0$ shows that λ is an eigenvalue of T.

Conversely, let $\lambda \in F$ be an eigenvalue of T. Let v be a non-zero vector in V such that $Tv = \lambda v$. Then

$$T^2 v = T(Tv) = T(\lambda v) = \lambda T(v) = \lambda . \lambda v = \lambda^2 v$$

Continuing this process, we have, $T^j v = \lambda^j v$ for each nonnegative integer j.

Thus,
$$\begin{aligned} o = p(T)v &= (a_0 + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m)v \\ &= (a_0 + a_1 \lambda + a_2 \lambda^2 + \cdots + a_{m-1} \lambda^{m-1} + \lambda^m)v \\ &= p(\lambda)v. \end{aligned}$$

Because, $v \neq 0$, the equation above implies that $p(\lambda) = 0$ which shows that λ is a zero of $p(z)$■

Thus, it is clear that the characteristic polynomial and the minimal polynomial of a linear operator T on a finite dimensional vector space have the same zeros. But do they have same number of zeros? The answer is negative. That is, if $\lambda \in F$ be the eigenvalue of a linear operator T on a finite dimensional vector space V then λ is a root of both characteristic equation and minimal polynomial equation of T. But if the multiplicity of λ in minimal polynomial equation of T is m and the multiplicity of λ in the characteristic equation of T is n then $m \leq n$.

Example. Let

$$A = \begin{pmatrix} 2 & -2 & 14 \\ 0 & 3 & -7 \\ 0 & 0 & 2 \end{pmatrix}$$

The characteristic polynomial of A, is given by

$$\det(xI - A) = \begin{vmatrix} x-2 & 2 & -14 \\ 0 & x-3 & 7 \\ 0 & 0 & x-2 \end{vmatrix} = (x-2)^2(x-3)$$

Since, the minimal polynomial and characteristic polynomial of A have same zeros, the only possibilities for the minimal polynomials are $(x-2)(x-3)$ or $(x-2)^2(x-3)$. Let us check now.

$$(A-2I)(A-3I)=\begin{pmatrix}0 & -2 & 14\\ 0 & 1 & -7\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}-1 & -2 & 14\\ 0 & 0 & 7\\ 0 & 0 & -1\end{pmatrix}=\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}.$$

Hence, the minimal polynomial is $(x-2)(x-3)=x^2-5x+6$.

Let us consider another example.

$$A=\begin{pmatrix}0 & -4 & 85\\ 1 & 4 & -30\\ 0 & 0 & 3\end{pmatrix}$$

Here the characteristic polynomial of A is given by

$$A=\det(xI-A)=\begin{vmatrix}x & 4 & -85\\ -1 & x-4 & 30\\ 0 & 0 & x-3\end{vmatrix}=x(x-4)(x-3)+4(x-3)$$

$$=(x-2)^2(x-3)$$

Since, the minimal and characteristic polynomials have same zeros, the possible minimal polynomials are $(x-2)(x-3)$ and $(x-2)^2(x-3)$.

Now, $$(A-2I)(A-3I)=\begin{pmatrix}-2 & -4 & 85\\ 1 & 2 & -30\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}-3 & -4 & 85\\ 1 & 1 & -30\\ 0 & 0 & 0\end{pmatrix}$$

$$\neq 0 \ [as\ (1{,}1) entry\ is\ non-zero]$$

So, $(x-2)(x-3)$ cannot be the minimal polynomial of A. Hence, the minimal polynomial of A is given by $(x-2)^2(x-3)$.

2.18 CANONICAL FORM

We know that it is easy to deal with diagonalizable linear operators defined on a finite dimensional vector space V. But we should remember that all linear operators are not diagonalizable. Thus we need to consider alternative matrix representations for non-diagonalizable linear operators. These representations are called *canonical forms.* There are different kinds of canonical forms depending on their applications. Here we shall discuss two types of canonical forms, *Jordan canonical form* and the *rational canonical form.*

The Jordan Canonical Form

Let T be a linear operator on a finite dimensional vector space V, and suppose that the characteristic polynomial of T splits, that is, the characteristic polynomial $f(z)$ of T can be written as in the form

$$f(z)=c(z-\lambda_1)(z-\lambda_2)\dots(z-\lambda_n)$$

where $\lambda_1,\lambda_2,..,\lambda_n$ are not necessarily distinct. If T is not diagonalizable, then there exists at least one λ_i whose geometric multiplicity is less than its algebraic multiplicity.

Better to study the concept in matrix form. We know that a square matrix A is similar to a square matrix B if there exists a non-singular matrix P such that $B=P^{-1}AP$. If A is diagonalizable, then we get B as $diag\ (\lambda_1,\lambda_2,..,\lambda_n)$. But what happens if A is not diagonalizable ? Definitely, we won't get a diagonal matrix similar to A but we will try to get nearly diagonal matrices. Now, we need to define some concepts.

A **Jordan block** J_i is a triangular matrix that has only one eigenvalue λ_i and only on eigenvector as given below

$$J_i=\begin{pmatrix}\lambda_i & 1 & ...0\\ 0 & \lambda_i & ...0\\ .. & .. & ..1\\ 0 & 0.. & \lambda_i\end{pmatrix}$$

where the principal diagonal is filled with some eigenvalue λ_i of the given matrix A, entries 1 on the diagonal next above the principal diagonal and all other entries are 0. (J_i may be just a 1-by-1 block consisting of just some eigenvalue).

For example, a 1×1 Jordan Block is given by (λ), a 2×2 Jordan Block is given by $\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix}$, a 3×3 order Jordan Block is given by $\begin{pmatrix}\lambda & 1 & 0\\ 0 & \lambda & 1\\ 0 & 0 & \lambda\end{pmatrix}$ and so on.

If a square matrix A of order n has s independent eigenvectors (if A is not diagonalizable then s must be less than n, i.e. $s<n$) then A is similar to a matrix J (**Jordan form**) with s blocks as given below

$$J=\begin{pmatrix}J_1 & 0 & ...0\\ 0 & J_2 & ...0\\ .. & .. & ..0\\ 0 & 0.. & J_s\end{pmatrix}$$

Where each J_i is a Jordan block and each 0 is a zero matrix.

For example, $J=$

$$\begin{pmatrix}2 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}$$

Here, we see that the eigenvalues of J are 2 with algebraic multiplicity 4 , 3 with algebraic multiplicity 2 and 0 with algebraic multiplicity 2. In other words, the

characteristic polynomial of J is $(x-2)^4(x-3)^2x^2$. It is clear from the matrix that the multiplicity of each eigenvalue is the number of times that the eigenvalue appears on the diagonal of J. Also observe that only 1st, 4th, 5th and 7th columns represent the eigenvectors of A.

But how to find a Jordan Canonical form of a given square matrix or a given linear operator on a vector space? No matter, whether the matrix or the operator is diagonalizable. Before that we wish to state a theorem without proof.

2.18.1 Theorem. Let A be an $n \times n$ matrix with characteristic polynomial

$$c(x) = (x-\lambda_1)^{k_1}(x-\lambda_2)^{k_2} \dots (x-\lambda_m)^{k_m}$$

and the minimal polynomial of A is given by

$$m(x) = (x-\lambda_1)^{t_1}(x-\lambda_2)^{t_2} \dots (x-\lambda_m)^{t_m}$$

then A is similar to a matrix J in Jordan Canonical Form

$$J = \begin{pmatrix} J_1 & 0 & \dots 0 \\ 0 & J_2 & \dots 0 \\ .. & .. & ..0 \\ 0 & 0.. & J_s \end{pmatrix}$$

where $J_1, J_2, .., J_s$ are all Jordan blocks. Furthermore, for each i

(i) the sum of the sizes of the Jordan blocks with diagonal entries λ_i is equal to k_i, that is, the algebraic multiplicity of λ_i.

(ii) the largest Jordan block with diagonal entries λ_i is of order $t_i \times t_i$ where t_i is the multiplicity of λ_i in the minimal polynomial $m(x)$.

(iii) the number of Jordan blocks with diagonal entries $\lambda_i$is equal to geometric multiplicity of λ_i.

Let us try to understand with an example. First of all I would like to clear two things. (1) Any square matrix can be brought into Jordan Canonical Form and (2) in the process characteristic polynomial and minimal polynomial of the given square matrix or of the given linear operator should be considered.

Let the characteristic polynomial $[c(x)]$ and minimal polynomial $[m(x)]$ of a given linear operator or of its associated matrix be as given below:

$$c(x) = (x-5)^3(x-3)^4$$
$$m(x) = (x-5)^2(x-3)^2$$

Thus it is clear from $c(x)$ that there are two eigenvalues, viz. 5 and 3 with algebraic multiplicity 3 and 4 respectively. We first consider the eigenvalue 5.

Now, look at $m(x)$. Here multiplicity of 5 is 2. Thus there exists a Jordan Block of order 2×2 corresponding to the eigenvalue 5 which is given below

$$J_1 = \begin{pmatrix} 5 & 1 \\ 0 & 5 \end{pmatrix}$$

There might be other Jordan Blocks corresponding to 5 but their order should be less than or equal to 2, i.e. either it is 1×1 block or 2×2 block. Again, by (i) of the above theorem, the sum of sizes of the Jordan Blocks should be equal to the algebraic multiplicity of 5, in this case it is 3. Since, size of J_1 is 2, other Jordan Block corresponding to 5 should be of size 1, that is,

$$J_2 = (5)$$

Now, consider the eigenvalue 3. It is clear from $c(x)$ that algebraic multiplicity of 3 is 4. Now, the minimal polynomial $m(x)$ ensures that there is Jordan Block of order 2 as given below

$$J_3 = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}.$$

Order of any other Jordan Block corresponding to 3 will be less equal to 2. In J_3, size of $3's$ is 2. But algebraic multiplicity of 3 is4. Hence, there are two possibilities. Either (i) a Jordan Block of size 2 corresponding to 3, as given below

$$J_4 = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$$

or (ii) there are two Jordan Blocks of size 1, viz. $J_5 = (3)$ and $J_6 = (3)$. Which possibility will be considered, depends on the geometric multiplicity of 3 as the total number of Jordan Blocks corresponding to the eigenvalue 3 equals to its geometric multiplicity. In this example, geometric multiplicity of an eigenvalue cannot be determined as the square matrix or the linear operator is not given. Thus, the required Jordan Canonical Form J, is given by either

$$J = \begin{pmatrix} J_1 & 0 & 0 & 0 \\ 0 & J_2 & 0 & 0 \\ 0 & 0 & J_3 & 0 \\ 0 & 0 & 0 & J_4 \end{pmatrix}$$

i.e.

$$\begin{pmatrix} 5 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

Or,

$$\begin{pmatrix} 5 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

2.19 RATIONAL CANONICAL FORM

If the characteristic polynomial of a square matrix or a linear operator splits then it is possible to bring it to Jordan canonical Form as described above. But what happens if the characteristic polynomial of a given square matrix or of a linear operator does not split, that is, if the square matrix has no eigenvalues in the given field F? For example, the characteristic polynomial of the matrix$A \in M_{2\times 2}(\mathbb{R})$, where $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is $x^2 + 1$ which has no eigenvalue in $\mathbb{R}$. In this case, rational canonical form needed. But what is rational canonical form? To answer this we must have an idea of companion matrix as described below.

Let us consider a monic polynomial $m(x)$ as

$$m(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1}x^{n-1} + x^n$$

Consider an$n \times n$ matrix $C(m)$ as given below

$$C(m) = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & \dots & 0 & -a_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & -a_{n-1} \end{pmatrix}$$

is a matrix such that the entries of the last column of $C(m)$ are filled by the coefficients of $m(x)$ (in increasing order of suffixes) with a negative sign, 1st$n-1$ diagonals are filled with 0 and entries below those $n-1$ diagonals are filled with 1 and all other entries are filled by 0. Then the matrix $C(m)$ is called **the companion matrix** of the monic polynomial

$$m(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1}x^{n-1} + x^n$$

For example, if we consider the monic polynomial of degree 3 as $m(x) = x^3 + 3x^2 - 8$.

Comparing with $a_0 + a_1 x + a_2 x^2 + x^3$ we see that $a_0 = -8$, $a_1 = 0$, $a_2 = 3$. Hence, the companion matrix $C(m)$ of $m(x)$ is given by

$$A = \begin{pmatrix} 0 & 0 & 8 \\ 1 & 0 & 0 \\ 0 & 1 & -3 \end{pmatrix}$$

Now, time has come to define *rational canonical form* of a given square matrix A.

It is mentioned here that for any $n \times n$ matrix A over a field F, there are uniquely determined monic polynomials $q_1(x), q_2(x), \dots, q_r(x)$ such that $q_1(x)|q_2(x)|\ldots q_{i-1}(x)|q_i(x) \ldots . q_{r-1}(x)|q_r(x)$ and $q_r(x)$ is the minimal polynomial of the matrix A. If $C_i(q_i)$ is the companion matrix of $q_i(x)$ for $i = 1,2,\ldots,n$ then the **rational canonical form** of A is the matrix with the block diagonal form

$$\begin{pmatrix} C_1(q_1) & 0 & \dots & 0 \\ 0 & C_2(q_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & C_r(q_r) \end{pmatrix}$$

We have deliberately avoided any reference to the underlying vector space andthe attendant relationship to the C_i's and invariant subspaces so as to achieve asimple description of the rational canonical form at least at the outset.

Remember the following things :

1. the elementary divisor $q_r(x)$ corresponding to $C_r(q_r)$ is the minimal polynomial of A
2. the product of the elementary divisors $q_i(x)$, $i = 1,2,\ldots,r$ is the characteristic polynomial $C(x)$ of A, that is ,
$$q_1(x).q_2(x) \ldots . q_r(x) = C(x) = \det(xI_n - A)$$
3. if$C(x)$ divides $q_r(x)$ then $C(x) = q_r(x)$, that is, in that case the characteristic polynomial of A and the minimal polynomial of A are the same.

Let us try with an example. Let

$$A = \begin{pmatrix} 2 & -2 & 14 \\ 0 & 3 & -7 \\ 0 & 0 & 2 \end{pmatrix}$$

The characteristic polynomial of A, is given by

$$\det(xI - A) = \begin{vmatrix} x-2 & 2 & -14 \\ 0 & x-3 & 7 \\ 0 & 0 & x-2 \end{vmatrix} = (x-2)^2(x-3)$$

Since, the minimal polynomial and characteristic polynomial of A have same zeros, the only possibilities for the minimal polynomials are $(x-2)(x-3)$ or $(x-2)^2(x-3)$. Let us check now.

$$(A-2I)(A-3I)=\begin{pmatrix}0&-2&14\\0&1&-7\\0&0&1\end{pmatrix}\begin{pmatrix}-1&-2&14\\0&0&7\\0&0&-1\end{pmatrix}=\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}.$$

Hence, the minimal polynomial is $(x-2)(x-3)=x^2-5x+6$.

Thus, the invariant factors are $(x-2)$and $(x-2)(x-3)$.

Companion matrix for $x-2$ is $C_1=(2)$

and companion matrix for $(x-2)(x-3)=x^2-5x+6$ is

$$C_2=\begin{pmatrix}0&-6\\1&5\end{pmatrix}.$$

Hence, the rational canonical form of A is

$$\begin{pmatrix}C_1&0\\0&C_2\end{pmatrix}$$

i.e
$$\begin{pmatrix}2&0&0\\0&0&-6\\0&1&5\end{pmatrix}\blacksquare$$

Let us consider another example.

$$A=\begin{pmatrix}0&-4&85\\1&4&-30\\0&0&3\end{pmatrix}$$

Here the characteristic polynomial of A is given by

$$A=\det(xI-A)=\begin{vmatrix}x&4&-85\\-1&x-4&30\\0&0&x-3\end{vmatrix}$$

$$=x(x-4)(x-3)+4(x-3)$$

$$=(x-2)^2(x-3)$$

Since, the minimal and characteristic polynomials have same zeros, the possible minimal polynomials are $(x-2)(x-3)$and $(x-2)^2(x-3)$.

Now, $$(A-2I)(A-3I)=\begin{pmatrix}-2&-4&85\\1&2&-30\\0&0&1\end{pmatrix}\begin{pmatrix}-3&-4&85\\1&1&-30\\0&0&0\end{pmatrix}$$

$$\neq 0\ [\,as\ (1{,}1)entry\ is\ non-zero]$$

So, $(x-2)(x-3)$ cannot be the minimal polynomial of A. Hence, the minimal polynomial of A is given by $(x-2)^2(x-3)$. Thus the only invariant factor is

$$(x-2)^2(x-3)=(x^2-4x+4)(x-3)=x^3-7x^2+16x-12.$$

Hence, the rational canonical form of A, i.e. the companion matrix of $x^3-7x^2+16x-12$, is given by

$$\begin{pmatrix}0&0&12\\1&0&-16\\0&1&7\end{pmatrix}\blacksquare$$

Solved examples :

1. **Find the minimal polynomial of the matrix A where A is given by**

$$A=\begin{pmatrix}3&0&1\\2&2&2\\-1&0&1\end{pmatrix}$$

Solution. The characteristic polynomial of A is given by

$$|xI-A|=\begin{vmatrix}x-3&0&1\\2&x-2&2\\-1&0&x-1\end{vmatrix}$$

$$=(x-3)(x-2)(x-1)+(x-2)$$

$$=(x-2)[(x-3)(x-1)+1]$$

$$=(x-2)^3$$

Since the minimal polynomial of A and the characteristic polynomial of A have same zeros, so possible forms of minimal polynomial of A are given by $(x-2)$ or $(x-2)^2$or $(x-2)^3$. Let us check.

Clearly, $A-2I\neq 0$. Now,

$$(A-2I)^2=\begin{pmatrix}1&0&1\\2&0&2\\-1&0&-1\end{pmatrix}\begin{pmatrix}1&0&1\\2&0&2\\-1&0&-1\end{pmatrix}=\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}.$$

Hence, the minimal polynomial of A is given by $(x-2)^2$.

2. **Let a, b and c be elements of a field F, and let A be the following 3×3 matrix over F :**

$$A=\begin{pmatrix}0&0&c\\1&0&b\\0&1&a\end{pmatrix}.$$

Prove that the characteristic polynomial for A is $x^3 - ax^2 - bx - c$ and that this is also the minimal polynomial for A.

Solution. The characteristic polynomial for A is given by

$$|xI - A| = \begin{vmatrix} x & 0 & -c \\ -1 & x & -b \\ 0 & -1 & x-a \end{vmatrix} = x\,(x^2 - ax - b) + 1(-c)$$
$$= x^3 - ax^2 - bx - c.$$

We first show that the minimal polynomial for A cannot be of degree 2.

Any monic polynomial of degree 2 can be written as $f(x) = x^2 + rx + s$ where $r, s \in F$.

Now,

$$f(A) = A^2 + rA + sI$$
$$= \begin{pmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{pmatrix}^2 + r\begin{pmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{pmatrix} + s\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 0 & c & ac \\ 0 & b & c+ba \\ 1 & a & b+a^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 & rc \\ r & 0 & rb \\ 0 & r & ra \end{pmatrix} + \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & s \end{pmatrix}$$
$$= \begin{pmatrix} s & c & ac+rc \\ r & b+s & c+ba+rb \\ 1 & a+r & b+a^2+ra+s \end{pmatrix}$$
$$\neq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus, $f(A) \neq 0$ for any $r, s \in F$. Hence, monic polynomial for A cannot be of degree 2. Therefore, the minimal polynomial for A must be of degree 3.

Since minimal polynomial for A divides $x^3 - ax^2 - bx - c$, the characteristic polynomial for A, the minimal polynomial for A is given by $x^3 - ax^2 - bx - c$.

3. Find a 3×3 matrix for which the minimal polynomial is x^2.

Solution. By the problem it is clear that any matrix A with $A \neq 0$ but $A^2 = 0$ serves our purpose. For example,

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

4. Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that there exists a polynomial $p \in \mathcal{P}(F)$ such that $T^{-1} = p(T)$.

Solution. Let $a_0 + a_1 z + a_2 z^2 + \dots + a_{m-1} z^{m-1} + z^m = 0 \dots (1)$ be the minimal polynomial of T, that is, this is the monic polynomial of smallest degree such that

$$a_0 I + a_1 T + a_2 T^2 + \dots + a_{m-1} T^{m-1} + T^m = 0 \dots (2).$$

If $a_0 = 0$, then multiplying both sides of (2) by T^{-1}, we get,

$$a_1 I + a_2 T + \dots + a_{m-1}(T)^{m-2} + T^{m-1} = 0$$

Therefore, we get a monic polynomial

$$q(z) = a_1 + a_2 z + \dots + a_{m-1} z^{m-2} + z^{m-1}$$

such that $q(T) = 0$ which contradicts that (1) is the minimal polynomial of T as degree of q is less than the degree of the polynomial given by (1). Thus, $a_0 \neq 0$.

Hence, by (2), we get

$$I = -\frac{a_1}{a_0} T - \frac{a_2}{a_0} T^2 - \dots - \frac{a_{m-1}}{a_0} T^{m-1} - \frac{1}{a_0} T^m$$

Operating both sides by T^{-1}, we get

$$T^{-1} = -\frac{a_1}{a_0} I - \frac{a_2}{a_0} T - \dots - \frac{a_{m-1}}{a_0} T^{m-2} - \frac{1}{a_0} T^{m-1}$$

Writing $p(z) = -\frac{a_1}{a_0} - \frac{a_2}{a_0} z - \dots - \frac{a_{m-1}}{a_0} z^{m-2} - \frac{1}{a_0} z^{m-1} \in \mathcal{P}(F)$

we have, $T^{-1} = p(T)$.

5. Give an example of an operator on $\mathbb{C}^3$ whose minimal polynomial equals z^2.

Solution. We wish to find an operator $T \in \mathcal{L}(\mathbb{C}^3)$ such that $T^2 = 0$.

Let us define $T : \mathbb{C}^3 \to \mathbb{C}^3$ by $T(z_1, z_2, z_3) = (z_3, 0, 0)$.

Then for $(z_1, z_2, z_3) \in \mathbb{C}^3$, we have, $T^2(z_1, z_2, z_3) = T(z_3, 0,0) = (0, 0, 0)$.

Thus, $T^2 = 0$.

If we take, $q(z) = z^2$, then we have, $q(T) = T^2 = 0$. Hence, minimal polynomial of T is a divisor of q. Now, the possible divisors of z^2 are $1, z, z^2$. Let $m(z)$ denote the minimal polynomial of T.

Now, $m(z) \neq 1$ as $m(T) = I \neq 0$.
If $m(z) = z$, then $m(T) = T \neq 0$.
So, $m(z)$ cannot be z.
Hence, $m(z) = z^2$.
Thus, the minimal polynomial of T, as given above, is z^2.

6. **Suppose $T \in \mathcal{L}(V)$ and $v \in V$. Let p be the monic polynomial of smallest degree such that $p(T)v = 0$. Prove that p divides the minimal polynomial of T.**

Solution. Let m denote the minimal polynomial of T. By division algorithm there exist polynomials $q, r \in \mathcal{P}(F)$ such that

$$m = qp + r, \quad 0 \leq \deg r < \deg p.$$

Thus, $m(T)v = q(T)p(T)v + r(T)v$.

But $m(T) = 0$ and $p(T)v = 0$. Thus, we have, $r(T)v = 0$.

Hence, $r = 0$(otherwise by a scalar multiplication r can be transformed to a monic polynomial r', such that $r'(T)v = 0$ and $\deg r' = \deg r < \deg p$ which is a contradiction).

Therefore, $m = qp$.

Thus, p divides the minimal polynomial of T.

7. **Find the Jordan Canonical form of the following matrix**

$$A = \begin{pmatrix} 1 & 5 & 7 \\ 0 & 4 & 3 \\ 0 & 0 & 1 \end{pmatrix}.$$

Solution. The characteristic polynomial of A is given by

$$\det(\lambda I_3 - A) = \begin{vmatrix} \lambda - 1 & -5 & -7 \\ 0 & \lambda - 4 & -3 \\ 0 & 0 & \lambda - 1 \end{vmatrix} = (\lambda - 1)^2(\lambda - 4).$$

Thus, the eigenvalues of A are given by 1and 4.

Since the characteristic polynomial of A and the minimal polynomial of A have same zeros, the possible minimal polynomials of A are given by $(\lambda - 1)(\lambda - 4)$and $(\lambda - 1)^2(\lambda - 4)$. Now,

$$(I - A)(4I - A) = \begin{pmatrix} 0 & -5 & -7 \\ 0 & -3 & -3 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & -5 & -7 \\ 0 & 0 & -3 \\ 0 & 0 & 3 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

So, we can rule out the possibility $(\lambda - 1)(\lambda - 4)$ as the minimal polynomial of A. Hence the minimal polynomial of A is given by $(\lambda - 1)^2(\lambda - 4)$.

Thus, for eigenvalue 1, there exists a 2×2 Jordan block as given by

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

and there is no other Jordan block corresponding to 1 as the algebraic multiplicity of 1 is 2. There is one 1×1 Jordan block corresponding to 4. Hence the Jordan Canonical form of A is given by

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

8. **Find the Jordan Canonical form of**

$$A = \begin{pmatrix} 2 & 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Solution. The characteristic polynomial of A is given by

$$\det(\lambda I_5 - A) = (\lambda - 2)^4(\lambda - 1).$$

[since determinant of an upper triangular matrix is the product of its diagonals].

Hence, the eigenvalues of A are given by 2 with algebraic multiplicity 4 and 1 with algebraic multiplicity 1. Since the minimal polynomial of A and the characteristic polynomial of A have same zeros, possible minimal polynomials of A are given below :

$m_1(\lambda) = (\lambda - 2)(\lambda - 1)$,

$m_2(\lambda) = (\lambda - 2)^2(\lambda - 1)$,

$m_3(\lambda) = (\lambda - 2)^3(\lambda - 1)$,

$m_4(\lambda) = (\lambda - 2)^4(\lambda - 1)$.

It is easy to show that $(A - 2I_5)(A - I_5) \neq 0$but $(A - 2I_5)^2(A - I_5) = 0$. Thus the minimal polynomial of A is $(\lambda - 2)^2(\lambda - 1)$. Therefore, we can say that there exists a 2×2 Jordan block corresponding to the eigenvalue 2, given by

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$$

and one1×1 Jordan block corresponding to the eigenvalue 1.

Let $E_2 = \{X : AX = 2X\}$.

Let $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} \in E_2$. Then, we have,

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ which gives } x_4 = 0,\ x_5 = 0.$$

Hence, $$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 0 \\ 0 \end{pmatrix} = x_1\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_2\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_3\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad x_1, x_2, x_3 \in \mathbb{R}.$$

Therefore, $E_2 = L(S)$ where

$$S = \left\{\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}\right\}.$$

Since, S is linearly independent, S is a basis for E_2. Hence, dimension of E_2 is 3, in other words, the geometric multiplicity of 2 is 3. Thus, there are three Jordan blocks corresponding to the eigenvalue 2, one is 2×2 and other two must be each of 1×1.

Hence, the Jordan Canonical form of A is given by

$$\begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

9. Find the rational canonical form of

$$A = \begin{pmatrix} 0 & -1 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}.$$

Solution. The characteristic polynomial of A is given by

$$\det(xI_3 - A) = \begin{vmatrix} x & 1 & 1 \\ 0 & x & 0 \\ 1 & 0 & x \end{vmatrix} = x^3 - x = x(x+1)(x-1)$$

Since the minimal polynomial of A and the characteristic polynomial of A have same zeros, the minimal polynomial of A is given by $x(x+1)(x-1) = x^3 - x$ which is same as characteristic polynomial of A and it is the only invariant factor. Hence, the rational canonical form of A is given by

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Exercise

1. Let V be a finite dimensional vector space over a field F. What is the minimal polynomial for the identity operator on ?

2. Let A be the 4×4 real matrix

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ -2 & -2 & 2 & 1 \\ 1 & 1 & -1 & 0 \end{pmatrix}$$

Show that the characteristic polynomial for A is $x^2(x-1)^2$ and it is also the minimal polynomial.

3. Find the Jordan Canonical form of

$$A = \begin{pmatrix} 3 & 0 & 0 & 2 & 0 \\ 0 & 3 & 0 & 0 & 5 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

4. Find the rational canonical form of

$$A = \begin{pmatrix} c & 0 & -1 \\ 0 & c & 1 \\ -1 & 1 & c \end{pmatrix}.$$